BULLETIN
OF THE
VIRGINIA POLYTECHNIC INSTITUTE

Engineering Experiment Station Series No. 149

PHYSICS OF THE SOLAR SYSTEM

PROCEEDINGS
OF THE CONFERENCE ON
PHYSICS OF THE SOLAR SYSTEM AND REENTRY
DYNAMICS
July 31 to August 11, 1961

Supported by a Grant
from the
NATIONAL SCIENCE FOUNDATION
and
Cosponsored by the
LANGLEY RESEARCH CENTER
of the
NATIONAL AERONAUTICS AND SPACE
ADMINISTRATION

Vol. LV No. 9 July, 1962

VIRGINIA ENGINEERING EXPERIMENT STATION

The Virginia Engineering Experiment Station was established in 1921 by action of the Board of Visitors of the Virginia Polytechnic Institute. It was established to stimulate and advance engineering education and to investigate problems of special importance to engineering, manufacturing, mining, transportation, and other industrial interests of the state and of the nation. Under the original arrangement, it was necessary for the departments of the school to carry on research within their normal operating budgets. This meant that monies used by each department for research under the program were used at the discretion of the departmental staffs and it also meant, therefore, that at most only meager amounts were available for research expenditures. Consequently, during the period between 1921 and 1950, research did not receive the impetus that it should have.

In 1950, the General Assembly of Virginia established the Virginia Engineering Experiment Station as a separate division of the Virginia Polytechnic Institute. Creation of the station as a distinct research organization within the school established the Station as an integral unit with its own budget. Benefits both to the institution and to industry have been multiplied many-fold since the 1950 legislation. The chief benefit, of course, has been to fulfill the chiefs aims, that is, to solve technical problems of industry and to extend knowledge; this has been accomplished through research carried on by staff members of the Engineering Experiment Station and the VPI School of Engineering and Architecture and also through short courses and "clinics" administered by members of these staffs. Of almost equal importance to the college as an educational institution, of course, has been the very real stimulation given to faculty and students through direct contact with and investigation of live problems in industry. Also noteworthy is the experience afforded research fellows who each year receive post graduate training and research know-how while carrying on investigations for the Station. These men gain experience that fits them especially well for important positions in industrial research organizations.

The Station Director is always happy to receive suggestions from industries regarding research and investigation. Where possible, means will be found for cooperation between the Station's investigators and individuals or organizations in conducting research. This is of special value to smaller industries which are individually unable to finance needed research. Commonly they will find that it is possible by group organization to support such work by the Experiment Station personnel at a nominal expense to each member of the group.

When research is undertaken for the private benefit of an industry, the industry will be expected to finance the work and reports will be made to those concerned. Results of investigations made at public expense and papers presented at the "clinics" will be published as bulletins of the Virginia Polytechnic Institute, Engineering Experiment Station Series, if they are considered to be of wide interest. Results of commercial value will be patented in accordance with contracts between the station and industries sponsoring the research. Typically, the proceeds are applied to promotion of the research work of the Station, with suitable recognition as to the rights of industries that may have assisted in financing the work.

The effects upon industrial development in Virginia of investigations now being carried on by staff members of the Virginia Engineering Experiment Station and VPI School of Engineering and Architecture will undoubtedly become readily apparent during the years ahead. It is hoped that as time goes on industry will realize more and more the value of this research. This will lead necessarily to more contributions by industry to the Station and to the School and will assure the full utilization of our research personnel and equipment.

A partial list of the Bulletins of the Virginia Polytechnic Institute, Engineering Experiment Station Series, which are still available, appears on the inside of the back cover. A complete list of available bulletins, including those of the Wood Research Laboratory Special Reports Series, can be obtained from the Director upon request.

John W. Whittemore, Director, Virginia Engineering Experiment Station, and Dean, VPI School of Engineering and Architecture

BULLETIN
OF THE
VIRGINIA POLYTECHNIC INSTITUTE

Engineering Experiment Station Series No. 149

PHYSICS OF THE SOLAR SYSTEM

PROCEEDINGS
OF THE CONFERENCE ON
PHYSICS OF THE SOLAR SYSTEM AND REENTRY
DYNAMICS
July 31 to August 11, 1961

Supported by a Grant
from the
NATIONAL SCIENCE FOUNDATION
and
Cosponsored by the
LANGLEY RESEARCH CENTER
of the
NATIONAL AERONAUTICS AND SPACE
ADMINISTRATION

Vol. LV No. 9 July, 1962

ACKNOWLEDGMENTS

VPI is deeply indebted to the National Science Foundation for its generous contribution in providing the funds for this Conference. This support was restricted almost entirely to travel expenses for the speakers and the one hundred faculty participants from American and Canadian Universities.

The help rendered by the Conference Steering Committee and their assistants, especially in the formative stages, was most welcome. This group was composed of Dr. Allen Waterman for NSF, Dr. John S. Stack for NASA, and Dean J. W. Whittemore for VPI.

The Langley Research Center of the National Aeronautics and Space Administration was likewise very helpful in assisting in the planning of this Conference and in contributing so generously by encouraging a number of its own personnel to take part in the program. Its help, especially that of Dr. Adolph Busemann, is gratefully acknowledged.

Thanks are also due a local Blacksburg electronics industry, The Polyscientific Corporation, for arranging an inspection of their plant showing the manufacture of slip rings used in many of America's satellites, and for sponsoring the reception hour just prior to the two banquets.

Lastly, the services of Professors Thomas E. Gilmer, Frank J. Maher, and others who assisted them in arranging for the living accommodations, local transportation, and conference meetings, and of Dr. Robert W. Truitt, former head of Aerospace Engineering and an original member of the Conference Committee through May, 1961, must be acknowledged with thanks.

The Conference Committee
J. B. EADES, Head, Aerospace Engineering
J. A. JACOBS, Head, Physics
D. H. PLETTA, Head, Engineering Mechanics, *Chairman*

PREFACE

Progress in basic and applied science and engineering has nowhere been as rapid as in the recent exploration of space. Although most of this advance would have been postponed materially if Federal funds had not been available, the fact remains that this research has been given top priority.

The present world political situation offers little hope that a lessening of our defense effort can be achieved. But even if lasting world peace were assured now, man's insatiable scientific curiosity would drive him on to exploring space beyond this earth. If the past has set any reliable precedent at all we should, then, expect the scientific discoveries which may result from such research and exploration to prove profitable as soon as they are exploited commercially. Indeed long-range weather forecasting and world-wide TV coverage are imminent. Hence, the drive to conquer space demands our best mental talent if either war or peace prevails.

Any nation's technical progress is related directly to the number of its talented people engaged in research and development. But before any latent talent can contribute much at the forefront of knowledge it must be developed, especially by training and education at the most advanced level.

The graduate engineers and scientists who staff the research and development laboratories of our defense effort, and the faculties of our technical schools, represent this nation's best talent. Although all are well trained, scientific progress is now expanding in almost exponential fashion, and these individuals can keep abreast of the new developments in their own specialties and allied fields only by attending periodic intensive conferences. Those who participate in such conferences are thus better able to carry on their research assignments or, if they teach, are better able to inspire their students. In their own environment they themselves serve as catalyzers.

In order to further this objective, the Virginia Polytechnic Institute (in cooperation with the Langley Research Center of the National Aeronautics and Space Administration) and the National Science Foundation cosponsored a two-week conference in the fields of Physics of the Solar System and Reentry Dynamics, from July 31 to August 11, 1961, inclusive.

Copies of most of the lectures were made available in manuscript form and are reproduced here. Abstracts are included in those cases for which complete manuscripts were not available. The information presented was either entirely new or else represented a digest of pertinent material that should prove useful.

These Proceedings are being issued in two parts for convenience. Part I includes those papers related to Physics of the Solar System. Part II includes those papers presented on Reentry Dynamics.

Page intentionally left blank

TABLE OF CONTENTS

CONFERENCE ATTENDEES

LIST OF SPEAKERS

KINSEY A. ANDERSON, *Asst. Prof.*
Department of Physics
University of California
Berkeley, California

ROGER A. ANDERSON, *Head*
Structural Configurations Branch
Langley Research Center, NASA
Langley Field, Virginia

GRANT ATHAY, *Acting Director*
High Altitude Obs.
Boulder, Colorado

CHARLES A. BARTH
Jet Propulsion Laboratory
Cal. Inst. of Technology
Pasadena, California

LUDWIG BIERMANN
Max Planck Institut fur Physik
und Astrophysik
Munchen, Germany

DEAN R. CHAPMAN
Aeronautical Research Scientist
NASA, Ames Research Center
Moffett Field, California

SYDNEY CHAPMAN
High Altitude Observatory
Boulder, Colorado

MACON C. ELLIS, JR.
Magnetoplasmadynamics Branch
Langley Research Center, NASA
Langley Field, Virginia

JOHN S. EVANS, JR.
Plasma Applications Section
Langley Research Center, NASA
Langley Field, Virginia

JOHN V. EVANS
Lincoln Laboratories
Mass. Inst. of Technology
Lexington, Massachusetts

ROBERT R. GILRUTH, *Director*
Space Task Group
Langley Research Center, NASA
Langley Field, Virginia

IRVING GRUNTFEST
System Engineering Department
General Electric Company
3198 Chestnut Street
Philadelphia 4, Pa.

RICHARD R. HELDENFELS, *Chief*
Structures Research Division
Langley Research Center, NASA
Langley Field, Virginia

RICHARD A. HORD
Dynamic Loads Division
Langley Research Center, NASA
Langley Field, Virginia

PAUL W. HUBER
Magnetoplasmadynamics Branch
Langley Research Center, NASA
Langley Field, Virginia

ABRAHAM HYATT, *Director*
Program Planning and Evaluation
NASA Headquarters
Washington, D. C.

E. S. LOVE, *Asst. Chief*
Aero-Physics Division
Langley Research Center, NASA
Langley Field, Virginia

WILLARD E. MEADOR, JR.
Space Physics Group
Langley Research Center, NASA
Langley Field, Virginia

CARL McILWAIN, *Research Asst. Prof.*
Department of Physics
State University of Iowa
Iowa City, Iowa

E. N. PARKER, *Prof.*
The Enrico Fermi Inst. for Nuclear
Studies
The University of Chicago
Chicago, Illinois

JOSEPH L. PAWSEY, *Asst. Chief*
Div. of Radiophysics
Radiophysics Laboratory
University Grounds
Sydney, Australia

GORDON H. PETTENGILL
Lincoln Laboratories
Mass. Inst. of Technology
Lexington, Massachusetts

LEONARD ROBERTS, *Head*
Mathematical Physics Branch
Theoretical Mechanics Division
Langley Research Center, NASA
Langley Field, Virginia

PETER H. ROSE
Avco-Everett Research Lab.
2385 Revere Beach Parkway
Everett, Massachusetts

GEORGE S. SCHAIRER, *Vice Pres.*
Research & Development
Boeing Aircraft, Seattle Division
Seattle, Washington

ALBERT H. SCHWICHTENBERG, *Head*
Dept. of Aerospace Medicine
Lovelace Foundation
4800 Gibson Blvd., S.E.
Albuquerque, New Mexico

SIEGFRIED F. SINGER, *Prof.*
Department of Physics
University of Maryland
College Park, Maryland

WILLIAM M. SINTON
Lowell Observatory
Flagstaff, Arizona

HARTLEY A. SOULE, *Asst. Director*
Langley Research Center, NASA
Langley Field, Virginia

JOHN WOOD
Smithsonian Astrophysical Obs.
Cambridge, Massachusetts

ROBERT B. VOAS, *Astronaut Training Officer*
Project Mercury
NASA Space Task Group
Langley Field, Virginia

BANQUET SPEAKERS

ROBERT C. SEAMANS, JR., *Assoc. Admin.*
NASA
Washington 25, D. C.

JOHN S. STACK, *Asst. Dir.*
Langley Research Center, NASA
Langley Field, Virginia

LIST OF PARTICIPANTS

DENYS O. AKHURST, *Head*
Dept. of Elect. Engr.
University of Arkansas
Fayetteville, Ark.

E. W. ANDERSON, *Head*
Aerospace Engr. Dept.
Iowa State University
Ames, Iowa

COLIN H. BARROW, *Visiting Assoc. Prof.*
Florida State University
Tallahassee, Florida

STANLEY BASHKIN, *Assoc. Prof.*
Physics Dept.
State University of Iowa
Iowa City, Iowa

R. W. BELL, *Professor*
Department of Aeronautics
U. S. Naval Post Graduate School
Monterey, Calif.

R. L. BOGGESS
E. E. Dept.
University of Michigan
Ann Arbor, Mich.

C. D. BOND, *Asst. Prof.*
Dept. of Physics
V. P. I.

MANFRED R. BOTTACCINI, *Prof.*
Aerospace Engr. Dept.
University of Arizona
Tucson, Arizona

MACK D. BOWEN
Engr. Dept.
Thompson Ramo Wooldridge Co.
Rocky Mount, Va.

T. G. BRNA, *Asst. Prof.*
Mechanical Engr. Dept.
V. P. I.

HARM BUNING, *Asst. Prof.*
Dept. of Aero. & Astro. Engr.
University of Michigan
Ann Arbor, Michigan

ADOLF BUSEMANN
NASA Langley Field
Langley Field, Virginia

D. L. BUSHNELL, *Asst. Prof.*
Dept. of Physics
V. P. I.

WINIFRED S. CAMERON
Theoretical Division, NASA
Silver Spring, Md.

FRANK CAREY, *Science Writer*
Associated Press
1300 Conn. Avenue
Washington 6, D. C.

PERCY H. CARR, *Prof.*
Physics Dept.
Iowa State University
Ames, Iowa

T. S. CHANG, *Assoc. Prof.*
Engr. Mech. Dept.
V. P. I.

LESLIE H. CHEN, *Head*
Applied Mechanics Section
Electric Boat Division
General Dynamics Corp.
Graton, Conn.

MARK H. CLARKSON, *Head*
Aerospace Engr. Dept.
University of Florida
Gainesville, Florida

CHARLES B. CLIETT, *Head*
Aeronautical Engr. Dept.
Mississippi State University
State College, Miss.

M. L. COLLIER, *Asst. Prof.*
Engr. Mechanics
V. P. I.

GAGE H. CROCKER, *Assoc. Prof.*
Dept. of Aeronautics
USAF Academy
Colorado

R. B. CROUCH, *Technical Asst. to U. Pres.*
Eastern Region
Lockheed Aircraft Corp.
1000 Conn. Avenue
Washington, D. C.

ELVIN J. DANTIN, *Assoc. Prof.*
Civil Engr. Dept.
Louisiana State University
Baton Rouge, Louisiana

FRED DE JARNETTE
Dept. of Physics
V. P. I.

JOHN P. DELVAILLE, *Res. Asst.*
Laboratory of Nuclear Studies
Cornell University
Ithaca, New York

B. L. DOW, *Chairman*
Aero. Engr. Dept.
Indiana Technical College
Fort Wayne, Indiana

RICHARD J. DRACHMAN, *Asst. Prof.*
Dept. of Physics
Brandeis University
Waltham, Massachusetts

J. B. EADES, JR., *Head*
Aero. Engr. Dept.
V. P. I.

HOWARD D. EDWARDS, *Res. Assoc. Prof.*
Dept. of Physics
Georgia Inst. of Technology
Atlanta 13, Georgia

GIOVANNI G. FAZIO
University of Rochester
Rochester 20, New York

H. KURT FORSTER, *Prof.*
Dept. of Engr.
University of California
Los Angeles 24, Calif.

MARK G. FOSTER, *Chairman*
Elec. Engr. Dept.
University of Virginia
Charlottesville, Virginia

DANIEL FREDERICK, *Prof.*
Engr. Mech. Dept.
V. P. I.

RICHARD GARRETT
Dept. of Physics
Hollins College, Roanoke, Va.

JOHN J. GIBBONS, *Prof.*
Dept. of Physics
Penn State University
State College, Pa.

DONALD A. GILBRECH
University of Arkansas
Fayetteville, Arkansas

THOMAS E. GILMER, JR., *Assoc. Prof.*
Dept. of Physics
V. P. I.

E. W. GLOSSBRENNER, *Engr.*
Poly-Scientific
Blacksburg, Virginia

W. J. GLOWACKI, *Aero. Engr.*
Naval Ordnance Lab. Maryland
Silver Spring, Maryland

JOHN W. GOAD, JR., *Engr.*
Thompson Ramo-Wooldridge
Rocky Mount, Va.

GEORGE GRAY, *Prof.*
Civil Engr. Dept.
V. P. I.

RAYMOND GRENCHIK, *Assoc. Prof*
Physics Dept.
Louisiana State University
Baton Rouge 3, Louisiana

GEORGE W. GRIMM
Ford Instrument Company
31-10 Thompson Avenue
Long Island City, New York

ANDREW GUTTMAN
Space Science Lab.
General Electric Co.
Philadelphia, Pa.

J. L. HAMMER, *Assoc. Prof.*
Civil Engr. Dept.
V. P. I.

DONALD E. HARDENBERGH, *Assoc. Prof.*
Engr. Mech. Dept.
The Pennsylvania State University
University Park, Pennsylvania

F. D. HARRIS, *Assoc. Prof.*
University of Missouri
Columbia, Missouri

ISADORE HARRIS, *Physicist*
Goddard Space Flight Center, NASA
Silver Spring, Md.

H. A. HASSEN, *Assoc. Prof.*
Aero. Engr. Dept.
V. P. I.

G. HEALY, *Asst. Prof.*
Engr. Mech. Dept.
V. P. I.

DONALD F. HEATH
Space Technology Center
General Electric Co.
Valley Forge, Pa.

S. M. HEFLIN, *Chairman*
Physics Dept.
Virginia Military Institute
Lexington, Virginia

SIDNEY HELFMAN, *Chief*
Fluid Dynamics Section
The Martin Company
Denver, Colorado

BOB HETERICK, *Asst. Prof.*
Civil Engr. Dept.
V. P. I.

ARTHUR HILL, *Reporter*
Roanoke Times
Roanoke, Virginia

JAMES A. JACOBS, *Head*
Physics Dept.
V. P. I.

WILLIAM T. JARRETT, *Group Engr.*
Research and Advanced Dev. Div.
AVCO
Wilmington, Mass.

THOMAS W. JACKSON, *Prof.*
Mech. Engr. Dept.
Georgia Institute of Technology
Atlanta 13, Georgia

J. N. JERKINS
Aero. Engr. Dept.
V. P. I.

PHILIP G. KIRMSER, *Prof.*
Ap. Mech. Dept.
Kansas State University
Manhattan, Kansas

F. M. KOBAYASHI, *Assoc. Prof.*
Engr. Sci. Dept.
University of Notre Dame
Notre Dame, Indiana

EDWIN H. KRUG, *Assoc. Mgr.*
Adv. Engr. Div.
Lear Inc.
Grand Rapids, Michigan

MARTEN LANDAHL, *Assoc. Prof.*
Massachusetts Inst. of Technology
Cambridge 39, Massachusetts

JOHN L. LANKFORD
Naval Ordnance Lab
Silver Spring, Md.

JAMES D. LAWRENCE, JR., *Asst. Prof.*
Physics Dept.
William and Mary College
Williamsburg, Va.

RICHARD S. LEE, *Asst. Prof.*
Mech. Engr. Dept.
N. C. State College
Raleigh, North Carolina

ALLEN M. LENCHEK, *Research Asst.*
Physics Dept.
University of Maryland
College Park, Md.

T. E. LEINHARDT, *Assoc. Prof.*
Physics Dept.
V. P. I.

R. G. LEONARD, *Instructor*
Mech. Engr. Dept.
V. P. I.

LAVERNE F. LEWIS, *Prof.*
Physics Dept.
School of Engineering AFIT
MCLI-W-PAFB, Ohio

CARL LILLIQUIST
Martin Company
Denver, Colorado

H. Y. LOH, *Prof.*
Physics Dept.
V. P. I.

DENNIS V. LONG, *Assoc. Prof.*
Engr. Mech. Dept.
V. P. I.

HSU LO, *Prof.*
Aero. & Engr. Sci. Dept.
Purdue University
Lafayette, Indiana

DAVID LUCK
RCA
Princeton, New Jersey

E. F. MAHER
Aeronautical Engr.
Naval Ordnance Lab.
Silver Spring, Md.

F. J. MAHER, *Professor*
Engineering Mechanics Dept.
V. P. I.

FRED W. MARTIN, *Prof.*
Aero. Engr. Dept.
Auburn University
Auburn, Ala.

R. W. MAJOR
Physics Dept.
V. P. I.

GEORGE E. MASE, *Assoc. Prof.*
App. Mech. Dept.
Michigan State University
East Lansing, Michigan

WILLARD E. MEADOR, JR.
Aerospace Technologist
Langley Research Center, NASA
Langley Field, Va.

M. A. MELEHY, *Assoc. Prof.*
Elec. Engr. Dept.
University of Connecticut
Storrs, Conn.

J. E. MILLER, *Prof.*
Physics Dept.
Clemson College
Clemson, South Carolina

R. H. MILLER, *Engr.*
Poly-Scientific Co.
Blacksburg, Va.

V. MADERSPACH, *Asst. Prof.*
Engr. Mech. Dept.
V. P. I.

HENRY MORRIS, *Prof.*
Civil Engr. Dept.
V. P. I.

VANCE E. MOYER, *Prof.*
Oceanography & Meteorology Dept.
A and M College of Texas
College Station, Texas

JOHN LACY MCKNIGHT, *Assoc. Prof.*
Physics Dept.
College of William and Mary
Williamsburg, Va.

BARRY J. NOONAN, *Aero. Res. Engr.*
U. S. Naval Ordnance Lab
Silver Spring, Md.

DANIEL B. OLFE, *Assoc. Prof.*
New York University
New York 53, New York

R. B. ORNDORFF, *Science Writer*
Richmond Times Dispatch
Richmond, Virginia

SIMON OSTRACH, *Prof.*
Mech. Engr. Dept.
Case Institute of Technology
University Circle, Cleveland, Ohio

GERALD A. OUELLETTE, *Lecturer, Supervisory Engr.*
Boston University & Raytheon Comp.
Boston, Mass.

A. PAP
Engr. Mech. Dept.
V. P. I.

HERMAN M. PARKER, *Prin. Scient. Rles.*
Aero. Engr. Dept.
University of Virginia
Charlottesville, Virginia

WILLIAM B. PARDO, *Asst. Prof.*
University of Miami
Coral Gables, Florida

V. L. PARSEGIAN, *Rensselaer Prof.*
Rensselaer Polytechnic Inst.
Troy, New York

RUSSELL E. PETERSEN, *Assoc. Prof.*
Mech. Engr. & Aero. Dept.
University of Arizona
Tucson, Arizona

GORDON H. PETTENGILL, *Assoc. Leader*
Lincoln Laboratory
Massachusetts Inst. of Technology
Cambridge, Mass.

D. H. PLETTA, *Head*
Engr. Mech. Dept.
V. P. I.

C. A. PLINT, *Assoc. Prof.*
Physics Dept.
University of Oklahoma
Norman, Oklahoma

WILLIAM E. READ, *Asst. Prof.*
USMA
West Point, New York

HARRY REAVES, *Instructor*
Physics Dept.
V. P. I.

DAVID D. REDFIELD, *Sci. Coord.*
Roanoke City Schools
Box 2129, Roanoke, Virginia

JOE W. REECE, *Instructor*
Engr. Mech. Dept.
N. C. State College
Raleigh, N. C.

ANDREW ROBESON, *Assoc. Prof.*
Physics Dept.
V. P. I.

SARA RICHARDSON, *Instructor*
Physics Dept.
V. P. I.

WEBB RICHARDSON, *Prof.*
Physics Dept.
V. P. I.

RICHARD RUSK, *Asst. Prof.*
Physics Dept.
V. P. I.

J. L. RYAN, *Asst. Prof.*
Physics Dept.
V. P. I.

S. T. STINSON, *Head*
Civil Engr. Dept.
Marshall University
Huntington 1, West Virginia

CHARLES H. SAMSON, JR., *Prof.*
Aero. and Civil Engr. Dept.
A and M College of Texas
Bryan, Texas

TURGUT SARPKAYA, *Assoc. Prof.*
University of Nebraska
Lincoln 8, Nebraska

WALTER J. SCHAETZLE
Mc Donnell A/C Corporation
St. Louis 66, Missouri

NORMAL SCHULTZ
University of Michigan
Ann Arbor, Mich.

REIN SILBERBERG, *Post-Doctoral Res. Assoc.*
U. S. Naval Research Lab
Washington, D. C.

WILLIAM M. SINTON, *Astronomer*
Lowell Observatory
Flagstaff, Arizona

A. WILEY SHERWOOD, *Prof.*
Aero. Engr. Dept.
University of Maryland
College Park, Md.

ROBERT SKULSKY, *Design Specialist*
Martin Co.
Denver, Colorado

ALEX G. SMITH, *Prof.*
Physics Dept.
University of Florida
Gainesville, Florida

C. W. SMITH, *Prof.*
Engr. Mech. Dept.
V. P. I.

DAVID P. STERN, *Research Assoc.*
NASA, University of Maryland
College Park, Md.

BERTRAM STILLER, *Physicist*
U. S. Naval Research Laboratory
Washington 25, D. C.

CONRAD D. SWANSON, *Chief*
Astronautical Engr. Section
NASA Marshall Space Flight Center
Huntsville, Alabama

G. SWIFT, *Asst. Prof.*
Engr. Mech. Dept.
V. P. I.

J. H. SWORD, *Assoc. Prof.*
Engr. Mech. Dept.
V. P. I.

WIDEN TABAKOFF, *Assoc. Prof.*
University of Cincinnati
Cincinnati 21, Ohio

J. B. TIEDEMANN, *Assoc. Prof.*
Aero. Engr. Dept.
University of Kansas
Lawrence, Kansas

JAMES H. TRAINOR, *Instructor*
University of New Hampshire
Durham, New Hampshire

TERRY TRIFFET, *Assoc. Prof.*
App. Mech. Dept.
Michigan State University
E. Lansing, Michigan

HUGH USSERY, *Prof.*
Physics Dept.
V. P. I.

GARVIN L. VON ESCHEN, *Prof.*
Aero. & Astro. Engr. Dept.
Ohio State University
Columbus 10, Ohio

JOHN A. WEESE, *Asst. Prof.*
Mech. Dept.
USAF Academy
Colorado

THOMAS C. WEST
Aero. Engr. Dept.
V. P. I.

R. K. WILL, *Assoc. Prof.*
Mech. Engr. Dept.
V. P. I.

DUDLEY WILLIAMS, *Prof.*
Physics Dept.
Ohio State University
Columbus, Ohio

ROGER WILLIS
Aero. Engr. Dept.
University of Princeton
Princeton, N. J.

J. GIBSON WINANS, *Assoc. Prof.*
Physics Dept.
University of Wisconsin
Madison 6, Wisconsin

JAMES T. YEN, *Asst. Prof.*
Mech. Engr. Dept.
N. C. State College
Raleigh, North Carolina

S. W. YUAN, *Prof.*
Aero. Engr. Dept.
The University of Texas
Austin 12, Texas

Page intentionally left blank

N63-10221

THE AMERICAN SPACE EFFORT

ROBERT C. SEAMANS, JR.
Associate Administrator,
National Aeronautics and Space Administration

I am most happy to meet with you at this Conference on Physics of the Solar System and Reentry Dynamics and to discuss the United States space effort. It is vital that leaders in basic and applied science—such as the members of this audience—understand why this country is expanding and accelerating its space exploration drive.

President Kennedy stated in his special message to Congress in May:

> "Now it is time to act, to take longer strides—time for a great new American enterprise—time for this nation to take a clearly leading role in space achievement."

Four major reasons underlie the national decision to marshal the resources required for leadership in space: 1) the quest for scientific knowledge; 2) direct and immediate application of satellites into operational systems; 3) the risk of delay in our competition with Communism; and 4) the technological advancement and stimulus to our economy that will emerge from the space effort.

Scientific Knowledge

Space research is a vigorously expanding field, whose growth is comparable to the development of nuclear physics after World War II. It is a field which cuts across the established areas of astronomy and physics and the earth sciences, and draws together scientists of varied backgrounds. The close interaction and exchange of ideas among scientists from many different fields have proven to be highly stimulating. Individuals from physics, geophysics, and astronomy have joined in attacking some of the most fascinating and significant problems to be found in modern research.

These are problems related to the manner in which the sun controls the atmosphere of the earth; to the structure of the earth; the moon, and the other bodies in the solar system; to the origin and history of the solar system; and to the structure and evolution of stars and galaxies. The scientists who work together on these investigations are united by their common interest in the world around them—the earth, its atmosphere, the solar system, the stars and galaxies. This is the spirit of the new science—an emphasis on the exploration of our environment.

The great volume of U. S. research in space science—approximately 120 space research articles published in a single U. S. journal, the Journal of Geophysical Research, during 1960—demonstrates the interest which American scientists have developed in this area. This high level of activity is asso-

ciated with rockets which can for the first time send major instruments, weighing several hundred pounds, into orbit above the atmosphere or out into interplanetary space. These rockets have opened up new avenues of attack on many important problems. Vital data are flowing into astronomy and the earth sciences, and revolutionizing these traditional disciplines.

Direct Applications

Space itself, when instrumented by man, will provide system capabilities not previously possible. Early returns from NASA experiments are already leading to early implementation of meteorological and communications satellite systems.

In 1960, NASA's Echo I passive communications satellite appealed to the world's imagination. The huge, aluminized plastic sphere has been seen like a moving star by people in many countries. Echo proved that it is possible to communicate between distant areas on the earth by reflecting radio signals from a satellite. Private companies have shown great interest in the Echo concept, and in "repeater" satellites that can receive and store messages at one point over the earth's surface and re-transmit them to ground receiving stations thousands of miles distant. Satellite communications will make world-wide telephone and television service realities, and will relieve the overload on present global communications systems. This enhanced communication could well be a bond drawing people of the world closer together.

Gen. David Sarnoff, chairman of the Radio Corporation of America, predicts:

> "Ten years hence there will be TV stations in virtually every nation on earth. An audience of a billion might then be watching the same program at the same time. The instrument which will give television's second epoch this distinctive global character is satellite relay transmission."

NASA's TIROS series of satellites has demonstrated the possibilities of vastly more accurate and longer-range weather forecasting. TIROS I transmitted 22,952 television pictures of the earth's cloud patterns. TIROS II, launched last November, reported important information about the atmosphere and the radiation of solar heat back from the earth.

Within 60 hours after the first TIROS was in orbit, its reports were being applied to day-to-day weather forecasting. In Hawaii, TIROS pictures helped trace the monsoons. Data on storms in the Indian Ocean were used by Australian meteorologists.

We are receiving excellent cloud pictures and infrared data from the weather satellite TIROS III which was launched July 12. The satellite was launched to coincide with the hurricane season, to gather information on the origin, development, and movement of these massive tropical storms. The Weather Bureau has made use of TIROS III pictures of Storm Eliza in the Pacific and Hurricane Anna in the Atlantic. NASA also used TIROS III for weather support of Astronaut Grissom's July 21 Mercury suborbital

flight. Twice a day as the satellite passed over the Caribbean, one of its two TV cameras was triggered to report weather conditions in the area of the flight.

According to the House Committee on Science and Astronautics, "An improvement of only 10 percent in accuracy (of weather forecasting) could result in savings totaling hundreds of millions of dollars annually to farmers, builders, airlines, shipping, the tourist trade, and many other enterprises."

Risk of Delay

It is not my place to discuss military missions but there is an important interchange of components and vehicles between the NASA and DOD programs. United States mastery of space is essential insurance against finding ourselves in two decades or less with a technology inferior to that the Russians will develop as they press forward on the space frontier. If we allowed them to surpass us, their space technology, in its military aspects, would be used to jeopardize our security.

In addition to potential direct military conflict, the Free Societies are in deadly competition with the Soviets for the support of the uncommitted peoples of the world. Space activity has great emotional appeal and we cannot afford the risk of being passed or appearing to be passed.

Today prestige is one of the most important elements of international relations. Essential to our prestige today is the belief of other nations that we have capability and determination to carry out whatever we declare seriously that we intend to do. There is no denying that in the eye of the world during the past few years our capability and determination have been brought into serious question.

In the minds of millions, dramatic space achievements have become today's symbol of tomorrow's scientific and technical supremacy. There is without a doubt a tendency to equate space and the future. Therefore, space is one of the fronts upon which President Kennedy and his Administration have chosen to act broadly, vigorously, and with continuous purpose. No other field offers us the opportunity to gain more of what we need abroad and at the same time achieve such a wealth of practical and useful results at home.

Stimulus to Economy

Our nation needs the stimulus, the knowledge, and the products that will evolve as we carry out our program to land Americans on the moon. The influence of the technical progress that will come into being through the integrating force and drive of a major space effort will be felt throughout the economy. Many of the instruments, equipment, power sources, and techniques that we devise to make manned lunar expeditions possible will be adaptable to other uses. The result will be great scientific advances and a variety of new consumer goods and industrial processes that will return tremendous benefits to us in practically every profession and activity.

Two decades ago the theme of the Temporary National Economic Committee hearings was that America's frontiers had closed and that this was

what had caused the stagnation of the thirties. The need of man's mind for new frontiers to cross has also been gravely concerning an increasing number of philosophers and psychologists. All frontiers seemed to have been passed, all new territories explored, with very little left for inquiring intelligence beyond applying and developing what has already been discovered.

Manned and unmanned exploration of space are already stimulating basic and applied research throughout our educational, government, and industrial systems. Now the concept of an eternally shut-in human race has been proved superficial. The prospect of exploring space is providing the catalyst and tonic for new adventures of the mind and spirit.

NASA Organization and Mission

As you know, the National Aeronautics and Space Administration directs U. S. space activities that are oriented toward scientific discovery and peaceful applications. Although in the public mind the space program overshadows our other work, we are also active in advanced aviation research in such fields as supersonic transport aircraft, vertical take-off and landing aircraft, safety, noise abatement, and special materials and structures.

At its birth nearly three years ago, NASA absorbed the National Advisory Committee for Aeronautics which for 43 years had been at the forefront of aviation research and which had also played an important role in early U. S. space research. To the 8,000 NACA scientists, engineers, and technical and administrative personnel utilizing five field centers, other excellent groups were added. Among these were the staff members of the Naval Research Laboratory Vanguard group which joined NASA in November 1958, and the approximately 2,400 people of the Jet Propulsion Laboratory, operated under NASA contract by the California Institute of Technology since December 3, 1958. On July 1 of last year, more than 5,000 staff members of the Development Operations Division, Army Ballistic Missile Agency, Huntsville, Alabama, were added. Today, our total employee strength—excluding Jet Propulsion Laboratory—is approximately 17,400, and by this time next year we plan on a strength of nearly 20,000.

We now have an all-around space research and development capacity. In this regard, may I emphasize that despite NASA's necessary growth during the last two years, we are determined that universities and industry shall get an ever-increasing share of NASA budget dollars. Contract participation currently amounts to more than 85 percent of those dollars. We feel that the NASA staff should be kept at a level necessary to plan the space exploration program and to organize, contract for and oversee it, while conducting enough in-house work to maintain the calibre of our scientific and technical personnel.

Of the $1,671,750,000 NASA budget for Fiscal Year 1962, $206,750,000 is for salaries and personnel expenses of the NASA organization. Contract effort provides for the construction of new facilities and the support of the research and development activities. The Fiscal Year 1962 budget includes $245,000,000 for construction of new and supporting facilities and $1,220,000,000 for research and development. NASA research and development

encompasses propulsion systems, propellants, power supplies, structures and materials, guidance and control, instrumentation and telemetry, and aerodynamics, as well as launch vehicles and the satellite program.

Space Sciences Program Highlights

Our space sciences program, which we carry out with satellites, space probes, and sounding rockets, consists of studies of the earth and its environment—geophysics—and of the sun, stars, and universe—astronomy.

Since January 31, 1958, this country has successfully orbited 45 earth satellites, two solar satellites, and two deep space probes. This program has provided important new scientific knowledge which has brought space into sharper focus and has contributed to the technology needed for more advanced spacecraft to come.

Some of the scientific findings are:

. . . Discovery of two intense radiation zones trapped around the earth—the Van Allen Belts.

. . . Determination that the earth is slightly pear-shaped with the stem at the North Pole. Other measurements by means of U. S. satellites have resulted in more accurate mapping of island and continent masses on earth.

. . . New data regarding the make-up of the fields of magnetism in space. For example, Explorer X, a 78-pound NASA satellite, transmitted highly meaningful data on solar-terrestrial relationships—such as magnetic fields and solar winds. Explorer X's data indicate that part of the interplanetary magnetic field near the earth may be an extension of the sun's magnetic field which has been carried earthward by solar winds.

. . . Discovery that sunlight exerts measurable physical pressure on objects in space. This pressure is shifting the orbit of the Vanguard I satellite about a mile per year and has affected the orbit of the 100-foot diameter Echo I satellite at a rate 300 times greater.

Among our most successful experiments to date have been the Pioneer series of space probes. Pioneer V, for example—launched into solar orbit on March 11 of last year—was tracked into space to a distance of 22.5 million miles, still the greatest distance any man-made object has been tracked. Pioneer V sent back scientific data on conditions in space until communication contact was lost on June 26, 1960. This space probe gave us new and valuable information about cosmic rays, the earth's magnetic field, solar "storms", and evidence of the existence of a large "ring current" circulating around the earth at altitudes of about 30,000 to 60,000 miles.

This gives you some idea of what we have been doing. In addition, our family of sounding rockets—which reach altitudes of from 150 to 4,000 miles—have been accumulating quantities of data on atmosphere composition and pressure, upper atmosphere winds, cloud cover, the make-up of the ionosphere, etc. And may I say, parenthetically, that during the next 12 months NASA will be launching more than 300 sounding rockets from Wallops Station, Virginia.

Advanced launch vehicles, such as the Agena and the Centaur, will soon be available to NASA. They will have greatly improved load-carrying capability. Detailed plans have been made and work has begun on an Orbiting Geophysical Observatory, based on the use of the Agena. This observatory will be one of our first standardized satellites, with a stock-model structure, basic power supply, attitude control, telemetry, and a command system. Its modular compartments are capable of carrying 50 different geophysical experiments on a single mission. The observatory itself will be about six feet long by three feet square. The two solar "paddles" which collect energy from the sun will be about six feet square. The satellite will weight 1,000 pounds and will include 150 pounds of scientific experiments.

NASA also has well-advanced plans for exploring the moon. A lunar spacecraft—known as Ranger—has been designed to carry an instrument package built ruggedly enough to survive a crash landing on the moon. Then its instruments will record and radio back to earth data on the make-up of the lunar surface. We plan to launch the first Ranger within the next few weeks.

Following Ranger will come Surveyor, a spacecraft that will be able to make a so-called "soft landing" on the moon. More delicate scientific instruments than those in Ranger can thus be employed.

Also under way is a spacecraft that will fly close to Venus and Mars, and later perhaps other, more distant planets. This spacecraft, called Mariner, will carry instruments to measure planetary atmosphere, surface temperatures, rotation rates, magnetic fields, and surrounding radiation regions. The Mariner series will be launched by our Centaur vehicle.

Each space flight has specific objectives. However, exploring space is like venturing along unknown streets in a strange town. We never know what is around the corner. Our early satellites discovered the existence of the Van Allen radiation around the earth. Undoubtedly future satellites will also discover phenomena which earthbound science could never deduce, much less observe. Judging by the experience of the past, the most intriguing—and perhaps the most valuable—discoveries in space exploration will be the unexpected ones.

Manned Space Flight

The United States has congratulated the Soviet Union on the orbital flights of Cosmonauts Gagarin and Titov. These achievements did not surprise us. We had been expecting them. Because the Russians have a five- or six-year lead on large boosters, we should be prepared for other Soviet "firsts" in space in the immediate future. This serves to underline the urgency of President Kennedy's decision to accelerate our own space program.

The historic flights of American Astronauts Alan Shepard and Virgil Grissom on May 5 and July 21, respectively, are being discussed at this conference by others more closely associated with the flights. As you know, these flights were important steps in Project Mercury, which is the first phase in the United States program for manned spaceflight. The primary objective of Mercury is to determine man's capabilities under the alien con-

ditions of acceleration and weightlessness with which future voyagers to the moon and planets will have to cope. These flights prove that man can perform useful tasks and can control a vehicle during spaceflight.

The spaceflights of Astronauts Shepard and Grissom were made to test the man and the Mercury spacecraft and to determine the quality of the vehicle and its systems and man's ability to handle them in space. In other words, the flights were made to learn how the astronaut, his capsule, and his equipment can best function together, as prefinal steps to putting an astronaut in orbit in space around the earth.

The first Mercury manned orbital flight will circle the earth three times and test man and capsule for about four and one-half hours. The Mercury capsule will return from the flight in a manner similar to the technique used in suborbital flights.

The second phase of our manned space flight program is called Apollo. The Apollo spacecraft will be large enough for living and working quarters to accommodate three men who will be able to operate in a "shirt-sleeves environment." It will be carried aloft by the Saturn booster, an eight-cluster rocket with a thrust of 1,500,000 pounds, compared to the Atlas with 360,000 pounds of thrust and the Redstone with only 78,000 pounds. The Redstone was used for the Shepard and Grissom flights and the Atlas will be the booster for Mercury orbital flights.

The Apollo-Saturn combination will first be used for an earth satellite, in which the three-man team can perform a great variety of tasks while training for sustained spaceflight. Next will come voyages deeper into space toward the moon and, in time, a three-man voyage around the moon and return to earth.

An actual moon landing is planned late in this decade. NASA's F-1 liquid-fuel rocket for the booster is now completed and ready for testing. This engine, developing 1.5-million pounds of thrust will be clustered into a booster called Nova.

We at NASA are sometimes asked why it is necessary to send men on hazardous spaceflights when instrumented satellites and probes are so versatile and have already returned such quantities of information on the near-space environment of the earth and on conditions in the vast reaches of deep space. Here, briefly, is my answer:

First, we already have an example of the need for manned participation in NASA's X-15 rocket airplane which has been flying to the fringes of space and has achieved a speed of 3,600 miles per hour. In at least eight out of thirty-eight X-15 flights to date, the flight would have failed without a pilot in the cockpit to correct malfunctions of equipment, instruments, or powerplant. In at least as many other cases, if X-15 missions had been unmanned, we would have obtained no information because either instruments or telemetry failed. The X-15 pilot, however, was able to land with valuable flight information recorded by his own senses.

Second, electronic instruments designed for NASA satellites and space probes can perform many tasks of sensing and measuring better than men could ever do. However, the statistical information gathered and transmitted

to earth by these instruments constitutes only a part of the basic research necessary for understanding the larger realities of space. The most advanced apparatus can perform only as it is programmed to do. Instruments have no flexibility to meet unforeseen situations. Scientific data acquired in space mechanically must be balanced by on-the-spot human senses, human reasoning, and by the power of judgment compounded of these human elements.

Dr. Carl Sagan of the Department of Astronomy, University of California, recently wrote to Senator Paul Douglas of Illinois to emphasize scientific reasons for manned spaceflight. I would like to pass on to you an excerpt from Dr. Sagan's letter:

"The scientific value (of spaceflight) comes when the men perform scientific tasks. There are large numbers of mineralogical, microbiological, and astronomical questions which trained scientific personnel on the moon will be able to answer far more reliably than any presently conceived automatic instruments . . .

"I feel strongly that, while an enormous amount of very significant scientific information can be obtained by unmanned vehicles, there are certain problems of the greatest significance which may well elude any unmanned system. If indigenous life exists on the planet Mars—and the bulk of contemporary evidence suggests that this is indeed the case—any but the most preliminary investigations will require a human experimenter.

"It is very difficult to imagine a sophisticated experimental program on the biochemistry, morphology, physiology, genetics, ecology or behavior of even simple extraterrestrial organisms carried out by a pre-programmed instrumented package. If the extraterrestrial organisms are very different from familiar lifeforms—and with five billion years of independent evolution, this may well be true—it is possible that an instrumented landing vehicle will not even be able to identify them as alive. A human scientist who can draw conclusions on the spot it an enormous asset in all aspects of lunar and planetary exploration.

Development of New Scientific and Technological Concepts

The use of manned and unmanned space vehicles provides the means by which university, industrial and government scientists can test many of their ideas and theories. In effect, the laboratory is extended into the universe. Here natural processes involving matter and energy are running their courses under extreme conditions and on enormous scales. Information which man-made devices can accumulate from these enormous laboratories is essential to scientific and technical progress.

In the development of a new concept in space science, the idea or hypothesis is suggested by a scientist as a result of his background of research and specialized knowledge. The proposal must then be subjected to a careful scrutiny by an appropriate group, including individuals from the various scientific and technological disciplines.

Once the objectives and the experimental approach have been thoroughly established, a project is initiated. Representatives from universities, Govern-

ment laboratories, and industry contribute to the design and laboratory perfection of the equipment and techniques for implementation of the experiment. Specifications for launch vehicles, spacecraft, and the necessary ground support and communication equipment are evolved. The necessary system components are procured or custom-developed as required. A flight operation is conducted to gather the data needed. Suitable data processing is carried out to provide experimental results in a form that can be compared with the original idea or hypothesis. This comparison may justify the original idea, or it may lead to acceptance in principle but modification in detail, or it may completely disqualify the concept.

Some concepts can be tested in conjunction with existing programs. Results can be achieved in a relatively short time, at moderate expense. Others are much more difficult and may require specially designed space vehicles traveling in specially calculated orbits. Several years may elapse between the suggestion of the idea and the availability of the relevant information from spaceflight research.

It can be seen that the space program requires the direct involvement of many highly qualified scientists, engineers, and technicians who understand how to act to accomplish the desired goals. However, we do not expect our schools, colleges, and universities to produce for us a specially educated and trained "space scientist or space engineer." We do not want to exploit so-called space science at the expense of weakening our national efforts in other scientific endeavors. The national program embracing space exploration and the associated technology is and should continue to be a part of a balanced national effort in *all science* and *technology*.

Exploration of Space—A National Goal

President Kennedy has said: "Over the next generation, there is no challenge facing our country that is more urgent or significant than that of providing better university education to more of our citizens. But it is a challenge we must face today—not a generation from now . . . We must have trained people—many trained people, their finest talents brought to the keenest edge. We must have not only scientists, mathematicians, and technicians. We must have people skilled in the humanities. For this is not only the age of the missile and space vehicle and thermonuclear power. This is the age that can become man's finest hour in his search for companionship and understanding and brotherhood."

Certainly, we all agree with this objective. The question arises whether the expansion of facilities, the addition of more scholarships, the introduction of new educational programs, important as they are, will suffice to implement these objectives. There is another all-important factor. There must be an inner desire on the part of our youth to fully utilize even the educational opportunities we have today. The feats of men are most potent in stimulating our youth to attain the heights of accomplishment for which education provides the capability.

These feats must be solid accomplishments of real and well-understood goals. What are the specific objectives of space exploration that satisfy this condition? It is imperative to continue our exciting scientific exploration of the atmosphere, the moon, the planets, the solar system, and the stars. It seems certain that the next decade will see general use of weather and communication satellites. Here virtually assured are early, practical, and extremely beneficial uses of space technology. These objectives represent great feats of science and engineering, but the greatest feat of all involves the flight of man himself in space. It must be remembered that man's entry into space is not a stunt, to be accomplished once, then discontinued. Rather, the program must broaden with new technology providing increasingly greater maneuverability and longer times of flight until man himself can explore the moon and planets.

The task of exploring space is one which will stretch the abilities and minds of men and women for years to come, providing a continuing long-term stimulant to our society. It will test to the utmost our powers to enlist the cooperation not only of teams of scientists and engineers but of every other element of our society. The magnitude of the task challenges the resources and cooperative will of all nations. We must all re-examine our habits of thought and action. We must use our insight and vision to guide our present decisions so that we can prepare our successors for the unfolding developments in the age of space exploration.

INTERPLANETARY DYNAMICAL PROCESSES: THE SOLAR WIND AND INTERPLANETARY FIELDS

E. N. PARKER

Enrico Fermi Institute for Nuclear Studies
and Department of Physics
University of Chicago

THE SOLAR WIND

Abstract

Hydrodynamic expansion of the solar corona is the basis for "solar corpuscular radiation". The quiet day coronal temperatures of $1 - 2 \times 10^6$ °K yield a solar wind of several hundred km/sec and a density of 5 — 50 particles/cm^3 at the orbit of Earth. The solar wind draws out into space the lines of force of the general one gauss solar field, to give an interplanetary field that is basically spiral in character with a density of the order of a few times 10^{-5} gauss at the orbit of Earth. The enhanced corona may sometimes have temperatures of 4×10^6 °K or more immediately following a large flare. Such explosive heating leads to a $1 - 2 \times 10^3$ km/sec blast wave into interplanetary space. The blast wave sweeps up the quiet-day solar wind ahead and kinks the quiet-day field, raising the density of the field to several times 10^{-4} gauss on some occasions. The galactic cosmic ray particles are swept back by the kink, producing a Forbush-type intensity decrease. Altogether, the blast wave and its magnetic field constitute the "enhanced solar corpuscular radiation".

It is shown that the alternative magnetic tongue model of Gold and others is untenable because the "tongue" would not reach the Earth until many hours after the arrival of the blast wave.

It is estimated that the supersonic solar wind extends into space to a distance of the order of 45 a.u. Stellar winds are expected to be a common phenomenon, and replace the Strömgren sphere for the cooler stars.

1 Introduction

Biermann pointed out some years ago (Biermann, 1951, 1952, 1957) that the acceleration and the excitation and ionization of type-I comet tails can be explained only by a continual high speed stream of particles from the sun; i.e., *solar corpuscular radiation.* His analysis showed that the corpuscular radiation is more intense during periods of enhanced solar activity, but that there is no indication that the radiation ever ceases, even during extended quiet periods. This view is consistent with the existence of the quiet-day aurora and the continual agitation of the polar geomagnetic field as solar corpuscular phenomena. Biermann originally estimated the cor-

puscular radiation flux to be of the order of $10^{10} - 10^{11}$ protons/cm^2-sec at the orbit of Earth, but more recently (Biermann, 1960) has revised his estimate to 10^9/cm^2-sec for the quiet-day value.

It is at once obvious that if Biermann is correct, then there can be no static interplanetary medium. All interplanetary gas would suffer the fate of the comet tail gas, being accelerated rapidly out of the solar system. On the basis of this we undertook to examine the existing static models of interplanetary space (e.g. Chapman, 1957, 1959). It soon became evident that the temperature of the solar corona (some $1 - 2 \times 10^6$ °K) is much too high to permit a static solar corona. The thermal conductivity of the coronal gas is so high as to extend the million degree temperature far into space, to such a distance that the thermal velocity far exceeds the gravitational escape velocity. Thus, to contain the solar corona in a static condition would require a strong confining pressure pushing in from interstellar space. The interstellar pressures are of the order of $10^{-14} - 10^{-12}$ dynes/cm^2 and are completely inadequate, by a factor of 10^8, to do the job. It follows, then, that the solar corona is not static but is expanding. The mean free path of the corona gas is small compared to the scale of the corona so that the appropriate treatment is with the hydrodynamic equation. Solution of the hydrodynamic equation showed immediately that the quiet-day corona must expand with a velocity reaching several hundred km/sec in space, and a total particle flux of $10^8 - 10^{10}$/cm^2-sec. This outward flow of coronal gas was termed the *solar wind* and is obviously to be identified with Biermann's continuing solar corpuscular radiation (Parker, 1958a).

Since this first identification of solar corpuscular radiation with the expanding corona, we have explored a number of consequences of the solar wind (Parker, 1958 b, c; 1960, 1961 a, b). It now appears that the solar wind is the dominant interplanetary dynamical force and is responsible for the interplanetary magnetic field configuration, the observed modulation of the galactic and solar cosmic ray intensity, the quiet day and the enhanced geomagnetic activity, etc.

There are many other phenomena, such as the aurora, geomagnetic activity, etc. which are conventionally ascribed to solar corpuscular radiation. They too are the consequence of the solar wind. On an interplanetary scale the solar wind is hydrodynamic, but on a planetary scale its low density renders it corpuscular in many respects.

Recent direct observation of the solar wind near the Earth (Bridge et al, 1961) has determined its quiet-day velocity to lie in the range 250-400 km/sec, with a density of 10-20 protons/cm^3. The total flux is of the order of 0.5×10^9 protons/cm^2-sec. An observation of the quiet-day solar wind in January and September of 1959 had been reported earlier by Shklovskii et al (1960) to suggest a flux of 0.2×10^9 protons/cm^2-sec, but the rather large corrections that had to be applied to the data make it uncertain as to how meaningful the flux estimate really was.

The remainder of this lecture is devoted to a discussion of the dynamical properties of the expanding corona and the solar wind, and to the inter-

planetary magnetic fields and cosmic ray variations which are the direct result.*

II. The Expanding Corona

A. Quiet Sun

The simplest model of the quiet corona is one which assumes spherical symmetry about the center of the sun. Then if $v(r)$ is the radial expansion velocity as a function of distance r from the center of the sun, the hydrodynamic equation for stationary expansion is

$$v\frac{dv}{dr} + \frac{1}{NM}\frac{dp}{dr} + \frac{GM_\theta}{r^2} = O \tag{1}$$

where N is the number of atoms per unit volume, G is the gravitational constant, and p is the hydrostatic pressure. Since the coronal gases are fully ionized and largely hydrogen, M is the mass of the hydrogen atoms and the hydrostatic pressure is $p = 2NkT$. Conservation of mass requires that

$$Nvr^2 = N_0v_0a^2 \tag{2}$$

where the subscript zero denotes the value at the reference level $r = a$. Choosing $a = 10^6$km, N_0 is of the order of $10^7/cm^3$. It will be sufficient for our purposes here to discuss the simple case of an isothermal corona. The more general case has been considered elsewhere (Parker, 1960).

For an isothermal corona, $T = T_0 \cong 2 \times 10^6$ °K, and (1) can be integrated to give

$$U^2 - \ln U - 2 \ln \zeta - H/\zeta = U_0^2 - \ln U_0^2 - H \tag{3}$$

upon using (2) to eliminate N. We have put $U = \frac{1}{2}\rho_0v_0^2/p_0$, $H = GM_\theta\rho_0/ap_0$, $\zeta = r/a$. The result is a one parameter family of curves $U(\zeta)$ for any given value of T_0, with U_0 as the parameter. The general form of the family is sketched in Fig. 1. The solution of physical interest is the solution starting from the origin and passing up through the critical point to supersonic velocity at infinity. The coordinates of the critical point are (U_c, ζ_c) where

$$U_c = 1/2^{\frac{1}{2}}, \quad \zeta_c = \tfrac{1}{2}H. \tag{4}$$

The value of U_0 for this solution is the lower positive root of

$$U_0^2 - \ln U_0 - H = -3/2 + \tfrac{1}{2}\ln 2 - 2 \ln H/2. \tag{5}$$

The coronal expansion will automatically drift into this solution whatever its initial expansion rate: The solutions with smaller U_0 go to $r = \infty$ on the lower branch of $v(r)$, which can be maintained only by a large inward

*This lecture is an extension of the material presented at the URSI Meeting, Washington, D. C., 3 April 1961 (to be published in Radio Propagation) and includes further results.

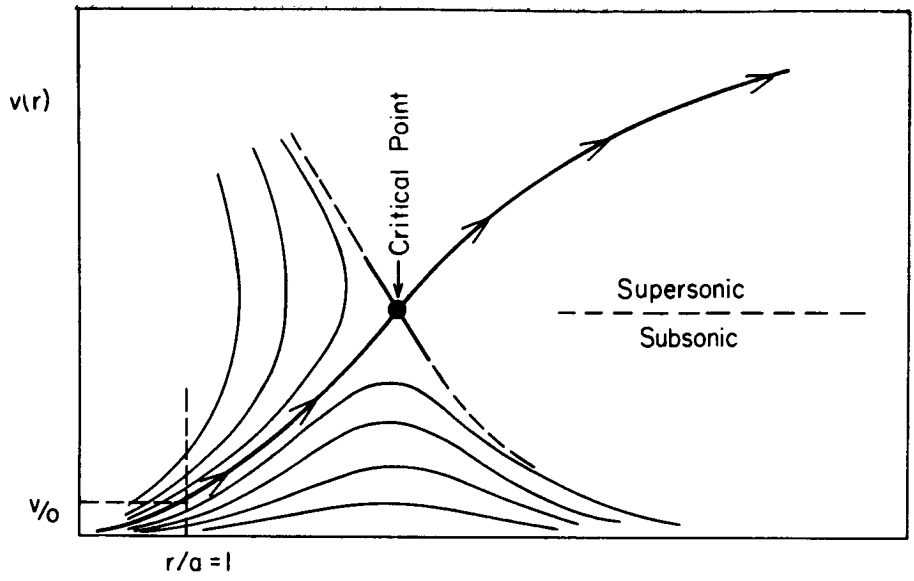


Fig. 1—The family of solutions of equation (3) for a given coronal temperature T_0. The solution of physical interest is the one passing through the critical point, given by equation (4). The critical point is analogous to the point of sonic transition in a Laval nozzle.

pressure from $r = \infty$. In the absence of such an inward pressure the expansion will accelerate to the solution through the critical point. There are no solutions above the solution through the critical point which start at the origin. Thus, the only available solution for expansion in the absence of an inward pressure at $r = \infty$ is the solution through the critical point, which starts with the velocity U_0 given by (5). It has been pointed out (Clauser, 1960) that this expansion of the corona is analogous to the expansion of a gas through a Laval nozzle, with the gravitational field playing the role of the throat.

The expansion velocity as a function of radial distance is shown in Fig. 2 for a number of coronal temperatures. The observed coronal temperatures are of the order of 2×10^6 °K (Billings, 1959), but even low temperatures like 1×10^6 °K give solar wind velocities of the order of several hundred km/sec.

A more general discussion of the expanding corona can be given on the basis of a polytrope law, in which the pressure p is related to the density ρ by $p = p_0(\rho/\rho_0)^\alpha$. The amount of heating of the corona by shock dissipation, thermal conduction, etc. is represented by the amount by which α is less than the adiabatic value 5/3. The velocity v(r) is then given by

$$U^2 + [\alpha/(\alpha-1)]\,(U_0/U\zeta^2)^{\alpha-1} - H/\zeta = U_0^2 + \alpha/(\alpha-1) - H. \qquad (6)$$

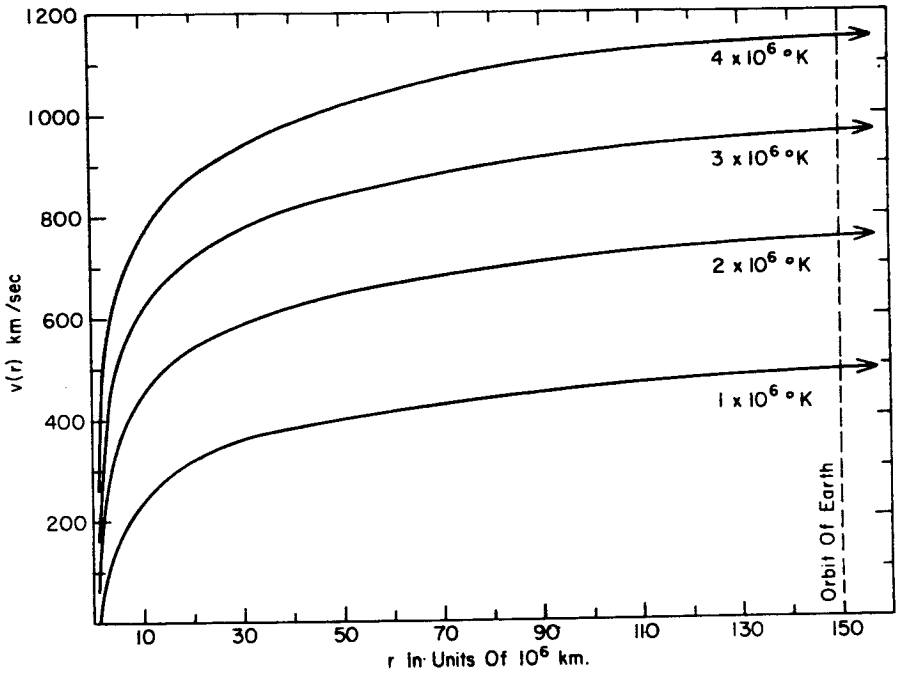


Fig. 2—The steady expansion velocity as a function of radial distance for a model isothermal corona. The two lower curves, for 1 and 2 x 10^6 °K, are applicable to the quiet corona. The upper curve, for 4 x 10^6 °K, gives an indication of the rate of expansion of the enhanced corona.

It is readily shown from this equation, and its differential form,

$$\frac{dU}{d\zeta}\left(2U - \frac{\alpha U_0^{\alpha-1}}{U^\alpha \zeta^{2(\alpha-1)}}\right) = \frac{2\alpha U_0^{\alpha-1}}{U^{\alpha-1}\zeta^{2\alpha-1}} - \frac{H}{\zeta^2}, \tag{7}$$

that a necessary and sufficient condition for supersonic expansion of the solar corona is $2\alpha < H < \alpha/(\alpha - 1)$. The restriction $H < \alpha/(\alpha - 1)$ is merely the requirement that the solar gravitational field not be so strong as to be able to contain the corona as a static atmosphere; the restriction $2\alpha < H$ is the requirement that the gravitational field be strong enough to simulate the throat of a Laval nozzle, permitting the transition to supersonic flow (at the critical point). Both conditions can be satisfied only if $\alpha < 3/2$. Numerically, $H \cong 8 \times 10^6/T$ for $a = 10^6$ km. The observed coronal temperatures of $1 - 2 \times 10^6$ °K give H in the range 4 — 8. On the other hand, observations of coronal temperature and density suggest that the effective value of α is in the vicinity of 1.1. It follows that the inequality is satisfied by the observed corona, leading to the conclusion that it expands supersonically.

Chamberlain (1960) has argued that for some reason the coronal temperature must have just such a value that $U_0^2 + \alpha/(\alpha - 1) - H = O$, leading to what he terms the *solar* breeze of some 18 km/sec. The internal inconsistency in the solar breeze model is exhibited in Chamberlain (1961) in which it is shown that the thermal conductivity of the solar corona conveys much too much energy outward to permit the slow solar breeze. Chamberlain postulates a reduction of the conductivity by transverse magnetic fields in the corona, contrary to the evidence from coronal streamers that the fields are principally radial. As a matter of fact the observed coronal temperature gradients, of the order of 3° K/km (Billings, 1960) predict a solar wind strength of about 3×10^8 protons/cm^2 sec, in rough agreement with observation, so that the conductivity is consistent with the observed solar wind, rather than the solar breeze. The calculation is rather simple: The thermal conductivity transports energy outward through the corona at a rate of the order of 2×10^{27} ergs/sec. There are not adequate radiative mechanisms in the outer corona to take care of this energy, so it must go into expansion. The lower limit of the expansion is then 2×10^{27} ergs/sec, which will lift 3×10^8 protons/cm^2 sec to the orbit of Earth against the solar gravitational field with a residual velocity of 300 km/sec upon arrival.

B. Active Sun

It is observed that following a large solar flare the coronal temperature over the general region of the flare may increase to 4×10^6 °K or more (Christiansen, Yabsley, and Mills, 1949; Christiansen and Warburton, 1953). The result is a hydrodynamic explosion of the corona outward into space. The expansion velocity may be in excess of 1000 km/sec, as is evident from Fig. 2. The asymptotic form of the resulting blast wave from the sun may be treated using the similarity transformations of the progressive wave (Courant and Friedrichs, 1949). The idealization is made that the corona has spherical symmetry, i.e. the entire corona is heated to 4×10^6 °K, rather than the portion over the active region in which the flare occurred. The 300 km/sec velocity of the quiet day solar wind ahead of the blast wave is neglected, and the temperature of the quiet day wind is assumed to be so small that the 10^3 km/sec velocity of the blast wave represents a very high Mach number. The resulting blast wave profiles are shown in Fig. 3. The shock transition at the front is of the collisionless type and involves a density increase by a factor of four above the quiet-day value. The shape of the blast wave behind the shock transition depends upon how hard the enhanced corona pushes from behind. In the event that the corona should push so hard that the energy of the blast wave increases linearly with time after leaving the sun, then the relatively thin, high-density profile identified by $\lambda = 1$ in Fig. 3 is the result. If the corona should not push at all once the blast wave is on its way, then a linear decrease of density behind the front is the result, identified by the sawtooth profile for $\lambda = 3/2$ in Fig. 3. In either case an enhanced coronal temperature of 4×10^6 °K near the

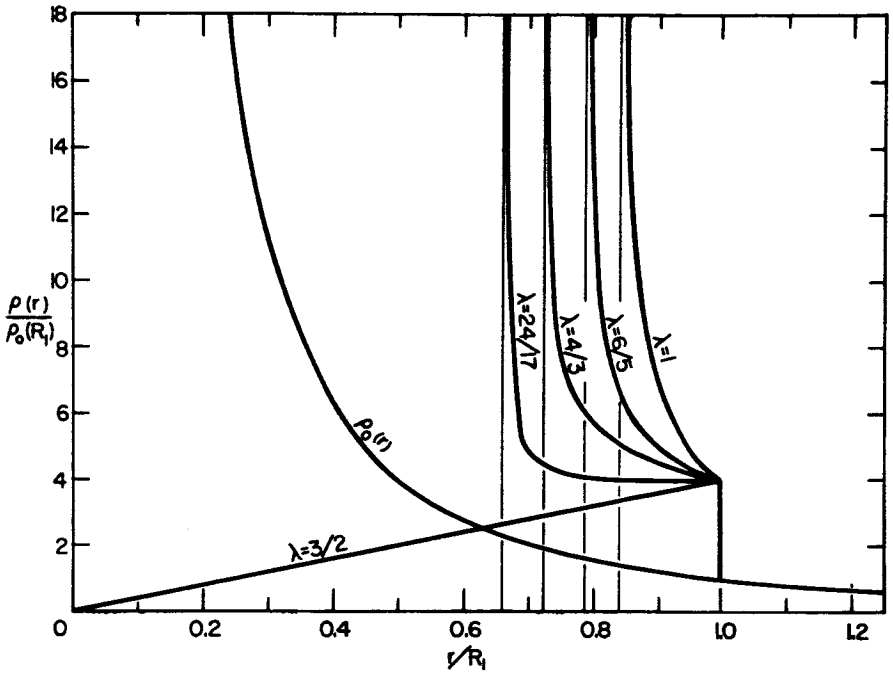


Fig. 3—Blast wave profiles from the active corona. The quiet day solar wind density is given by ρ_0 (r), and is essentially proportional to $1/r^2$. The shock transition at the head of the blast wave is at $r = R_1$, where the density jumps by a factor of four from the quiet day value. The parameter λ is an inverse measure of how strongly the enhanced corona is driving the blast wave from the rear: If the corona does not push at all on the blast wave, then $\lambda = 3/2$ and the sawtooth profile is the result; if the corona pushes so hard that the energy of the blast wave increases linearly with time, then $\lambda = 1$ and the blast wave lies between $r/R_1 = 0.84$ and $r/R_1 = 1.0$.

sun yields a blast wave with a velocity of 1-2 $\times$ 10^3 km/sec, and a density of the order of 10^2/cm^3 at the orbit of Earth (Parker, 1961). Thus the blast wave represents the "enhanced solar corpuscular radiation" responsible for the geomagnetic storm, the Forbush decrease, etc.

An important point to be noted is that the blast wave consists of gas which had already left the sun before the occurrence of the flare. It is material which has been swept up as the wave advances outward into space. The material in the corona at the time of the flare is behind the rear of the blast wave profile. This must be borne in mind in the next section when we discuss the interplanetary magnetic field configuration.

III. Interplanetary Magnetic Fields

A. Quiet Sun

The magnetic fields in interplanetary space are composed of the magnetic lines of force of the general solar field (Babcock and Babcock, 1955;

Babcock, 1959) which are pulled outward from the sun by the expanding corona. The magnetic fields associated with active regions on the sun are usually considerably in excess of one gauss, so that they can to a large degree withstand the expansion of the local corona, remaining fixed over the active regions. The hydrostatic pressure in the solar corona is of such an order, 10^{-2} dynes/cm^2, as to suggest that only fields of the order of one gauss can be extended by the solar wind. The simplest model, then, is one in which a corona with spherical symmetry expands outward, carrying with it the dipole-like fields of the polar regions of the sun. Such an idealized picture ignores the many complications of the disordered one gauss fields which are undoubtedly carried outward from the equatorial regions. The idealization also ignores the many interesting complications that result from the fact that the actual corona expands more rapidly and densely in some directions than in others. However, the idealization of a radial solar wind with spherical symmetry in a dipole field serves to illustrate the general principles and to give a rough idea of the overall interplanetary magnetic configuration to be expected. When more comprehensive observations allow greater detail to be filled into the present simple theoretical models, the many irregularities and swirls which undoubtedly exist in the interplanetary field may be taken into account. Ignoring the irregularities now will give a conservative estimate of the cosmic ray effects to be discussed in the next chapter.

The lines of force of a general solar dipole field represent the two parameter family of curves.

$$r - a \cong \frac{v}{\Omega} (\phi - \phi_0), \ \theta = \theta_0$$

upon extension into space by the quiet day solar wind of velocity v. The solar angular velocity Ω is approximately 2.6×10^{-6} radians/sec, and (θ_0, ϕ_0) is the position, in spherical polar coordinates, of the intersection of the line of force with the reference level $r = a$ in the lower corona. The radical component of the field is

$$B_r (r, \theta) = B_r (a, \theta) \left(\frac{a}{r}\right)^2,$$

the meridional component is zero, and the azimuthal component is

$$B_\phi (r, \theta) = B_r (r, \theta) \, r\Omega/v.$$

Projection of each line of force onto the solar equatorial plane is an Archimedes spiral, shown in Fig. 4 for $v = 300$ km/sec. It is readily seen that the field is principally radial inside the orbit of Earth and principally azimuthal beyond. One gauss at the solar photosphere yields about 2×10^{-5} gauss in the radial direction at a distance of one astronomical unit.

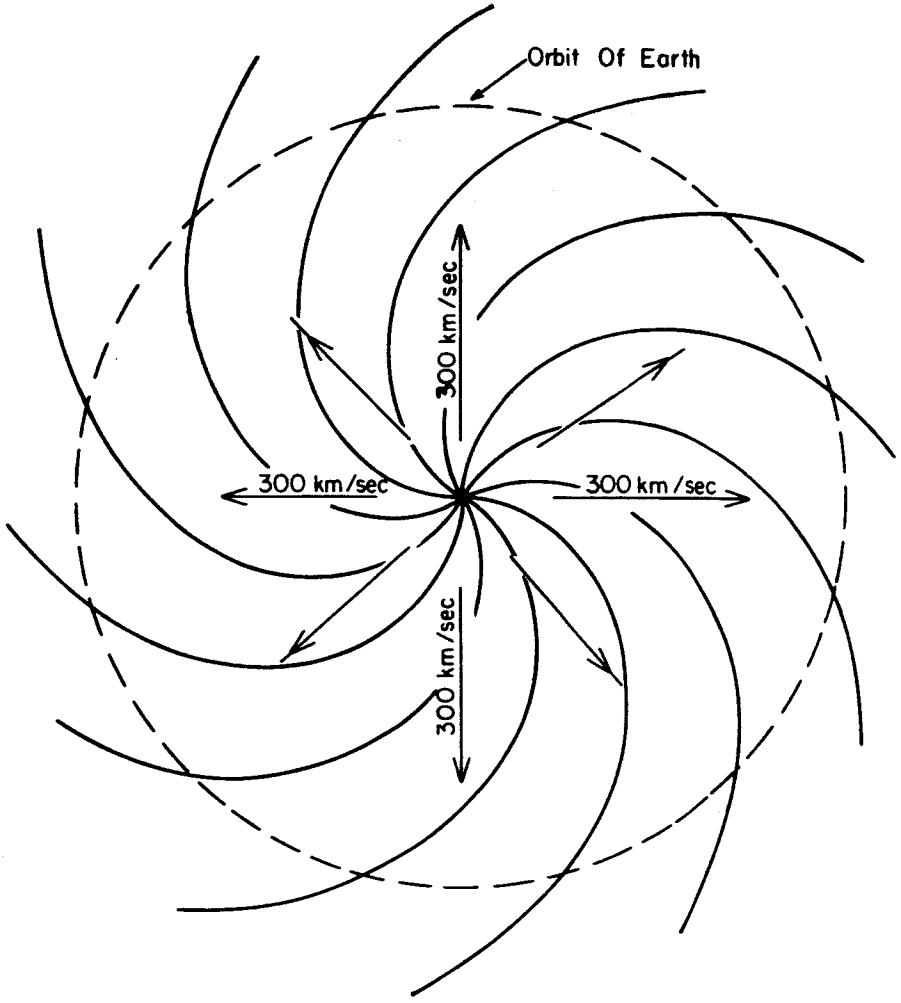


Fig. 4—Projection onto the equatorial plane of the lines of force of the solar fields extended by a quiet-day radial solad wind of 300 km/sec.

B. Active Sun

The quiet-day interplanetary field shown in Fig. 4 is deformed by the blast wave from the corona over a flare as shown in Fig. 5. Again the case $\lambda = 1$ represents the situation when the enhanced corona is pushing hard on the rear of the blast wave, and $\lambda = 3/2$ when the blast wave is coasting. In either case the field may reach 10×10^{-5} gauss at the front of the wave. When the corona is pushing hard on the rear of the blast wave the field toward the rear may easily be 40×10^{-5} gauss or more.

It has been suggested that the "enhanced solar corpuscular radiation", which we suggest is a hydrodynamic blast wave from the corona, some-

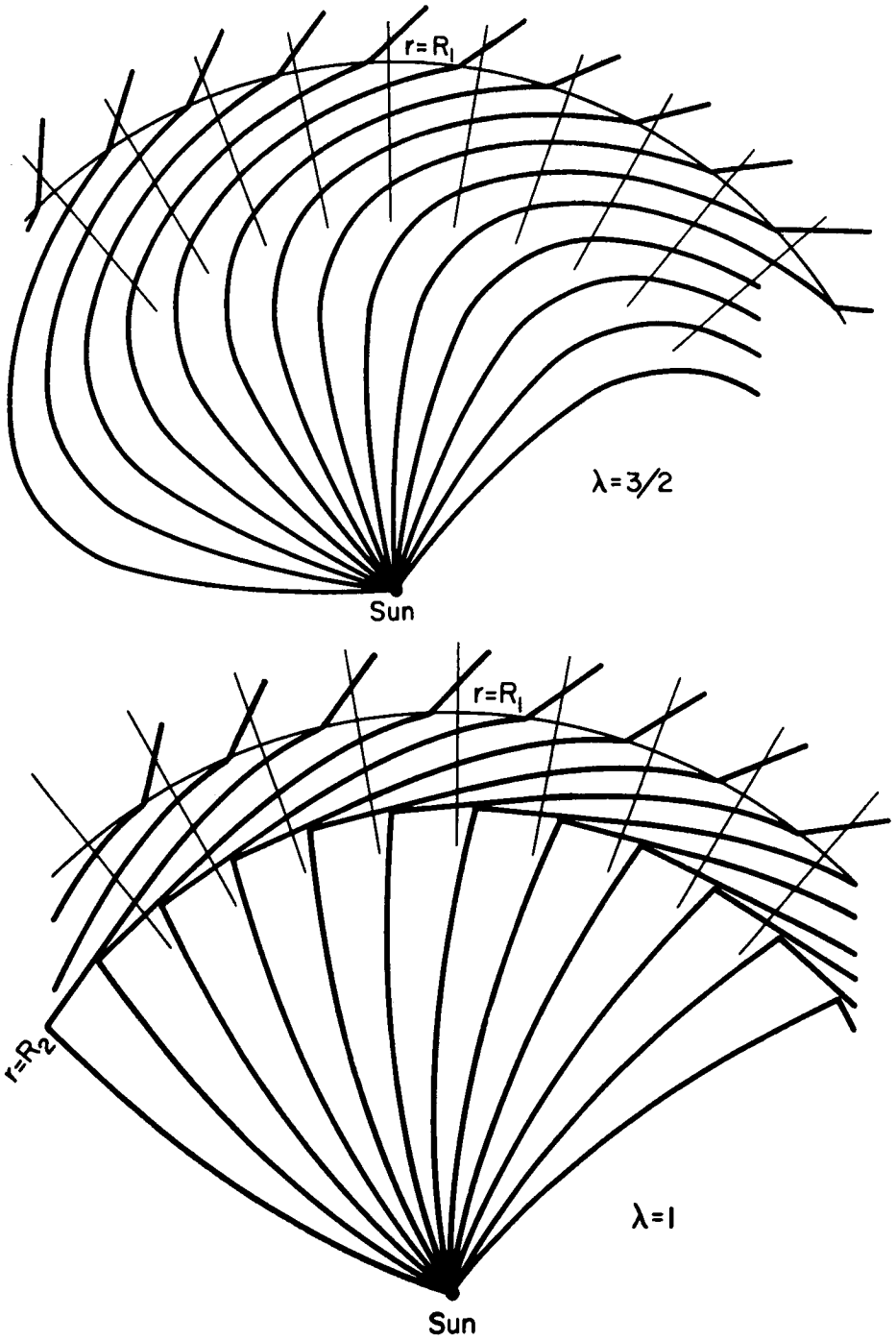


Fig. 5a and b—Deformation of the quiet-day field shown in Fig. 4 by the freely coasting blast wave, $\lambda = 3/2$, and by the driven blast wave $\lambda = 1$. The shock transition is at $r = R_1$. The rear of the blast wave is at $r = R_2$.

times draws out a loop of field from the active region, giving a magnetic tongue which is responsible for many of the effects associated with the enhanced corpuscular radiation. (see discussion and references in Gold, 1960). Presumably the expanding interior of the magnetic tongue is partially shielded by the fields of the tongue from the galactic cosmic ray intensity, giving the Forbush-type cosmic ray decrease in coincidence with the geomagnetic storm resulting from the impact of the gas in the tongue against the geomagnetic field; once the tongue has engulfed Earth, it affords free access of solar protons from the sun to Earth; the tongue stores the solar protons following the emission from the sun, etc. In this way it has been argued that the tongue explains most of the observed cosmic ray phenomena associated with "enhanced solar corpuscular radiation."

We object to the tongue model of the interplanetary field on a number of points. First of all, the magnetic configuration of the blast wave fields shown in Fig. 5 possess the same properties as the tongue so far as modulating and channeling galactic cosmic rays and solar protons. Thus the tongue is not unique and seems an unnecessary complication. Second, and more serious, is the objection that the tongue cannot possibly arrive at Earth with the "enhanced solar corpuscular radiation". Whatever the nature of the outburst on the sun which leads to the enhanced radiation, the result is a blast wave which scoops up the interplanetary gas ahead to form a wave as shown in Fig. 3. The active coronal material is behind the blast wave, and hence so is the magnetic tongue which the coronal material is supposed to carry. Even in the extreme case that the enhanced corona is pushing so hard on the rear of the blast wave that $\lambda = 1$ (which requires that the 10^6 °K coronal temperatures extend all the way to the rear of the wave) the tongue is 0.16 a.u. behind the front of the blast wave, which means a delay of the order of 6 hours in the arrival of the tongue. Yet it is clear in many cases that the cosmic ray modulation etc. begins with the first onset of geomagnetic activity. The tongue configuration is shown in Fig. 6 for the special case that $\lambda = 3/2$. The tongue extends barely beyond the orbit of Mercury when the blast wave reaches the orbit of Earth.

IV. Cosmic Ray Modulation

A. Quiet Sun

The general outward flow of nonradial magnetic fields from the sun tends to push back the galactic cosmic ray particles and thereby decrease the cosmic ray intensity in interplanetary space below the interstellar level. This reduction follows the general solar cycle and is usually referred to as the 11-year variation. There are many individual effects which probably contribute to the total reduction: The idealized quiet-day interplanetary field shown in Fig. 4 is, in effect, a giant corkscrew conveyor which revolves once every 27 days with the sun and tends to push outward the galactic particles sliding inward along the lines of force. There is reason to expect disordering in the interplanetary field, some examples of which have been cited elsewhere (Meyer, Parker, and Simpson, 1956; Parker,

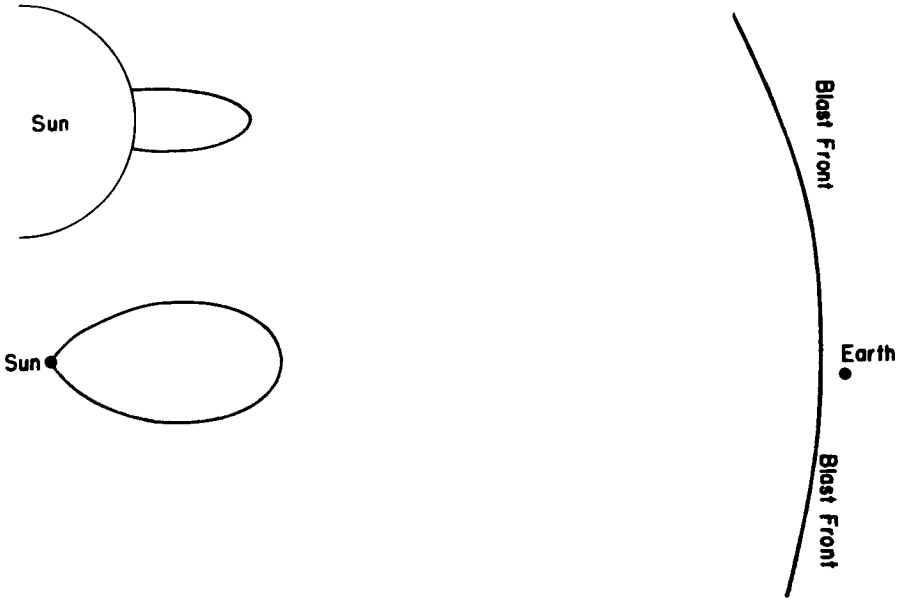


Fig. 6—A re-entrant loop of magnetic field in the solar corona before an outburst is shown, followed by its extension into space by the blast wave $\lambda = 3/2$. The loop remains in association with the coronal gas with which it moves following the outburst. The blast wave ahead of the loop consists of interplanetary gas (the quiet-day solar wind) swept up to form the blast wave.

1958c); the disordering contributes to the trapping of solar particles and to the impediment of galactic cosmic ray particles diffusing inward from interstellar space.

These effects are large and contribute significantly to depression of the cosmic ray intensity. The depression is of the order of a factor of three at 1 Bev (see for instance Simpson, 1960) and probably more at lower energies. We may hope that one day interplanetary observation may decide what effects predominate.

B. Active Sun

The magnetic configuration carried by the blast waves from the active corona, and shown in Fig. 5 for an idealized case, have the basic characteristic of a strong magnetic field (probably 10-50 $\times$ 10^{-5} gauss) in the wave, with an essentially radial field connecting into the sun from behind. The strong fields result from the twisting and pinching together of the quiet-day lines of force. The resulting constriction in the lines of force represents a serious impediment to the passage of cosmic ray particles. Thus the blast waves will have a tendency to store energetic protons of solar origin behind them, and to push back galactic cosmic ray particles ahead of them, as required by observation (see for instance, Steljes, et al, 1961).

The Forbush decrease for the idealized model shown in Fig. 5 may be as large as 40 per cent, and more if any complications are permitted in the idealized field configuration. The energy dependence of the cosmic ray decrease can be estimated if it is remembered that the blast waves are in fact not spherical but occupy a solid angle of the order of one steradian. The diffusion of cosmic rays from the sides of the region swept out by the blast waves gives an energy dependence to the fractional decrease of the cosmic ray intensity as shown in Fig. 7. The parameter ν in the figure is a measure of the angular width of the blast wave. The computed energy dependence is found to be not unlike the observed energy dependence of the Forbush decrease (see for instance MacDonald and Webber, 1960) viz. proportional to reciprocal rigidity, or flatter.

V. Stellar Winds

Having presented a brief outline of the solar wind, and the interplanetary magnetic fields and cosmic ray effects which result, consider the phenomenon in a somewhat larger perspective. The sun is evidently a normal class G5 main sequence star in the middle of its life. The sun has a million degree

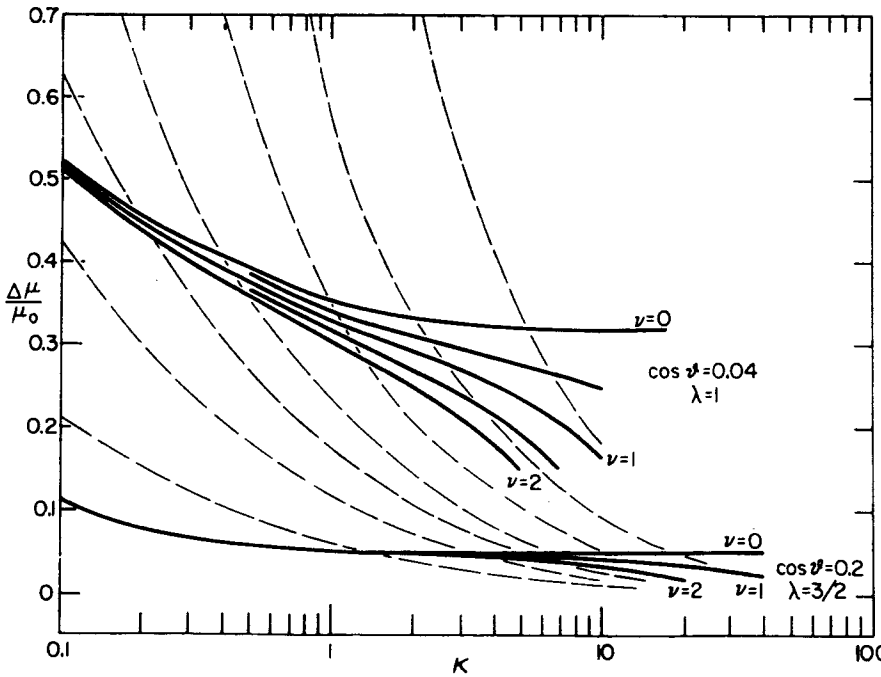


Fig. 7—The fractional depression $\Delta\mu/\mu$ of the cosmic ray intensity as a function of proton energy κ (in units of 931 Mev) for a freely coasting shock, $\lambda = 3/2$, and for the hard driven shock, $\lambda = 1$. The parameter ν is an inverse measure of the angular width of the blast wave as seen from the sun. The broken lines represent reciprocal rigidity, for purposes of comparison.

corona which expands to give the supersonic solar wind of some 300 km/sec. The solar wind appears to be a normal feature of a normal period in the life of the sun. Hence, it is not unreasonable to suppose that there may have been a solar wind ever since the sun reached the main sequence 5×10^9 years ago. It is not unreasonable to suppose that there will continue to be a solar wind at least until the sun leaves the main sequence some 10^{10} years hence. By the same token, one would suppose that most other class G main sequence stars in the galaxy possess stellar winds. In fact it would seem plausible that most, if not all, other stars with extended coronas must have some sort of stellar wind, because of the enormous outward energy transport by the thermal conductivity of coronal gas. Schatzman (1960) has pointed out that one expects stars cooler than about the middle of class F to have deep convective zones. Since it is presumably acoustical waves, hydromagnetic waves, etc. generated in the convective zone which heat the corona, we might reasonably speculate that most stars cooler than class F have extended coronas and stellar winds (Parker, 1960).

The stellar wind of a star pushes back the interstellar gas and fields until some sort of momentum balance is reached between the wind pressure $\frac{1}{2}$ NMv^2 and the interstellar gas and field pressure p. The wind pressure, decreases outward from the star like $1/r^2$. For a 300 km/sec solar wind of 10 protons/cm^3 at one astronomical unit, the wind pressure is 0.8×10^{-8} dynes/cm^2. The interstellar pressure of a 10^{-5} gauss interstellar field is about 4×10^{-12} dynes/cm^2, suggesting that the supersonic solar wind may extend to some 45 a.u., i.e. the orbit of Pluto. If this distance is typical, then the stellar wind regions of the cooler main sequence stars do not occupy a large portion of interstellar space (45 a.u. is only about 2×10^{-4} of the distance to the nearest star), but on the other hand they replace the Strömgren sphere* as the region of ionized hydrogen around the cooler stars.

The hydrodynamics of the stellar wind region as a whole is of some interest and can be worked out in idealized cases fairly simply (Parker, 1961b). The supersonic stellar wind undergoes a shock transition to subsonic flow at approximately the radial distance where $\frac{1}{2}$ NMv^2 is equal to the interstellar pressure. The subsonic flow beyond the shock transition is shown in Fig. 8, for the case of a stellar wind with spherical symmetry in the presence of an interstellar wind, i.e. a motion of the star relative to the interstellar medium. If a large-scale interstellar magnetic field is present, the subsonic flow inflates a cavity in the field, which may be anything from a cylinder to a sphere, depending upon the stagnation pressure Π^2 of the gas flow at infinity. The shape of the cavity is shown in Fig. 9 relative to the shock surface for various values of Π^2, where Π^2 is the stagnation pressure of the subsonic flow (determined by conditions at infinity) measured in units of the undisturbed interstellar field pressure.

*The Strömgren sphere is the spherical region of hydrogen around a star kept ionized by the ultraviolet light of the star (Strömgren, 1939, 1948). The radius of the sphere, the *Strömgren radius*, may be of the order of parsecs for an O-star, but diminishes very rapidly, like $T^{1/3}\ 10^{-2280/T}$, with the surface temperature T and is only of the order of astronomical units for the cooler stars.

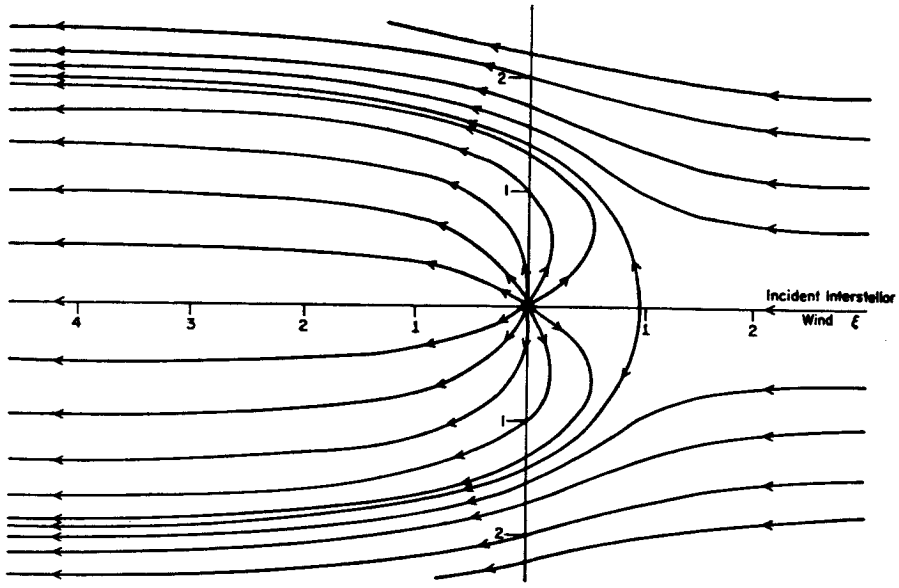


Fig. 8—The streamlines of the subsonic hydrodynamic flow of a stellar wind beyond the shock transition where $\frac{1}{2}NM\,v^2 \cong P$, in in the presence of a subsonic interstellar wind carrying no significant magnetic field.

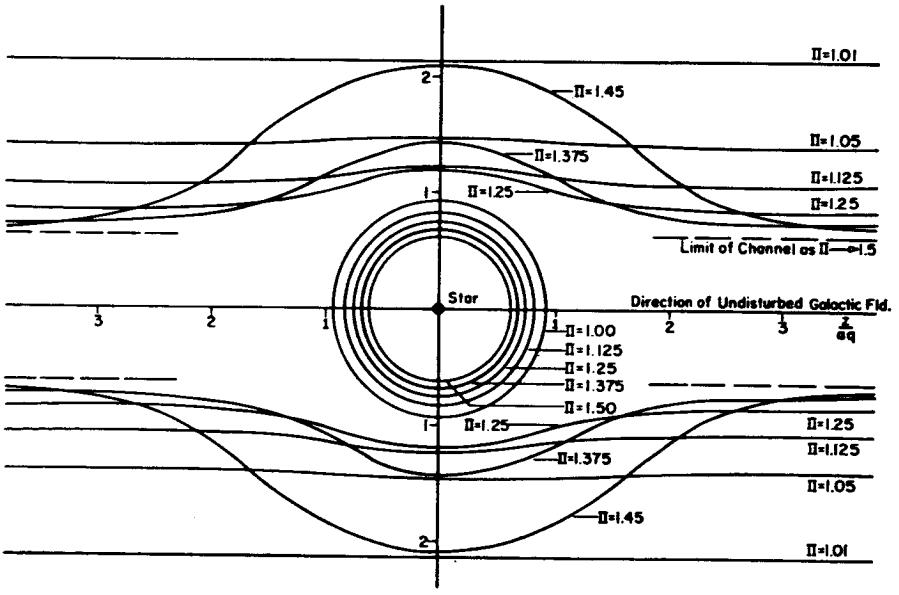


Fig. 9—The shock transition to subsonic flow, shown by the concentric circles, and the outer boundary of the stellar wind region in the presence of a large scale interstellar magnetic field, for various values of the stagnation pressure π^2.

VI. Summary

The purpose of the paper has been to describe in a semi-quantitative way the solar wind model of interplanetary dynamical processes. The model affords a deductive approach to the problem of interplanetary plasmas, fields, and cosmic ray variations, starting with the hydrodynamic theory of the expanding solar corona. The observed coronal temperatures and their variations lead through such idealized models as we have discussed to the kind of "corpuscular" and cosmic ray effects that are known from observation. As direct observation of interplanetary conditions become more comprehensive with the progress of space experimentation, we look forward to filling in the many details which are missing from the present simple model.

References

Babcock, H. D., 1959, *Astrophys. J.* **130,** 364.
Babcock, H. W. and Babcock, H. D., 1955, *Astrophys. J.* **121,** 349.
Biermann, L., 1951, *Zeit. f. Astrophys.* **29,** 274.
....., 1952, *Zeit, f. Naturforsch.* **7a,** 127.
....., 1957, *Observatory,* **107,** 109.
....., 1960, 4th Symposium on Cosmical Gas Dynamics, Varenna, Italy (August).
Billings, D. E., 1959, *Astrophys. J.* **130,** 961.
....., 1960, Private Communication.
Bridge, H., Lazarus, A., Lyon, C. F., Rossi, B., and Scherb, F., 1961, Spring Meeting of AGU, Washington, D. C., April.
Chamberlain, J. W., 1960, *Astrophys. J.* **131,** 47.
....., 1961, *ibid,* **133,** 675.
Chapman, S., 1957, *Smithsonian Contr. Astrophys.* **2,** 1.
....., 1959, *Proc. R. Soc. London, A.* **253,** 462.
Christiansen, W. N. and Warburton, J. A., 1953, *Australian J. Phys.* **6,** 190.
Christiansen, W. N., Yabsley, D. E., and Mills, B. Y., 1949, *Australian J. Sci. Res. A* **2,** 506.
Clauser, F. H., 1960, 4th Symposium on Cosmical Gas Dynamics, Varenna, Italy (August).
Courant, R. and Friedrichs, K. O., 1948, *Supersonic Flow and Shock Waves,* New York, Interscience Publishers, pp. 419-421.
Gold, T. H., 1960, *Suppl. Astrophys. J.* **4,** 406.
MacDonald, F. B. and Webber, W. R., 1960, *J. Geophys. Res.* **65,** 767.
Parker, E. N., 1958a, *Astrophys. J.* **128,** 664.
....., 1958b, *Physics of Fluids,* **1,** 171.
....., 1958c, *Phys. Rev.* **110,** 1445.
....., 1960, *Astrophys. J.* **132,** 821.
....., 1961a, *Astrophys. J.* **133,** 821.
....., 1961b, *Astrophys. J.* **134** (July issue).
Schatzman, A., 1960, *Mt. Wilson Pub.*
Shklovskii, I. S., Moroz, V. I., and Kurt, V. G., 1960, *Astron. Zh.* **37,** 931.
Strömgren, B., 1939, *Astrophys. J.* **89,** 526.
....., 1948, *Astrophys. J.* **108,** 242.
Simpson, J. A., 1960, *Suppl. Astrophys. J.* **4,** 378.
Steljes, J. F., Carmichael, H., and McCracken, K. G., 1961, *J. Geophys. Res.* **66,** 1363.

STRUCTURE OF THE SOLAR ATMOSPHERE

R. GRANT ATHAY
Acting Director, High Altitude Observatory
Boulder, Colorado

Abstract

The structure of the sun's outer layers, including the photosphere, chromosphere, and corona, and the phenomena of solar activity are discussed from the standpoint of their thermodynamic properties and the physical processes giving rise to them. Emphasis is placed on the major unsolved problems and uncertainties existing in the science of solar physics. It is noted that many of the standard techniques used for analysis of the sun's atmosphere represent gross over-simplifications of the actual problem. Promising new theoretical and observational approaches to these problems, and the role of the sun in relationship to space research are reviewed.

SOLAR RADIO ASTRONOMY

JOSEPH L. PAWSEY
Radiophysics Division C.S.I.R.O.
Sydney, Australia

1. Introduction

In these two lectures I propose to give a general account of the observation of the Sun using radio waves, and what such observations, combined of course with pertinent optical ones, are telling us of physical phenomena in the Sun.

The Sun is a great ball of very hot gas, but, viewed optically, it appears to have an abrupt boundary—the photosphere. Above this boundary the atmosphere is transparent; we can see nothing below it. The reason is that the density increases progressively downwards and with it the opacity. At the level of the photosphere, the opacity to light happens to increase rather quickly. A similar phenomenon occurs with radio waves, but the "visibility limit" rises with increasing wavelength from the mid-chromosphere at one centimetre, to the base of the corona at about one metre, and to a height of about one solar radius at about 10 metres. These limits of visibility are indicated in Figure 1. In the radio case the major cause of opacity is one familar to all ionospheric physicists. The outer atmosphere of the Sun is almost completely ionized and for each wavelength the level at which the electron density is the critical density for the wavelength concerned is a limiting level below which the refractive index is imaginary and waves do not propagate. These levels, of course, rise with increasing wavelength, since the density decreases with height.

Thus, using radio waves, we expect to be able to "see" features of the solar atmosphere above the visibility level for the wavelength concerned, provided these features are suitably illuminated. In principle, the illumination may arise from self emission by the features concerned, from illumination by a terrestrial transitter (radar), or from an external cosmic source (scattering of waves from radio stars). Of these, the self emission case dominates the picture and my lectures deal almost exclusively with solar radio emission. However a certain amount of information about structure in the outer corona has been derived by the scattering method and I shall touch on this. The possibility of obtaining solar echoes was pointed out by Kerr in 1952, and the first echoes actually obtained by Eshleman, Barthle and Gallagher in 1960. But as yet these studies have not contributed to our knowledge of the Sun and I shall not discuss the radar method further.

The region in the Sun accessible to radio observations is thus the chomosphere and the corona, and in this region there are a wide variety of disturbances ranging from persistent clouds of excess density to prominences

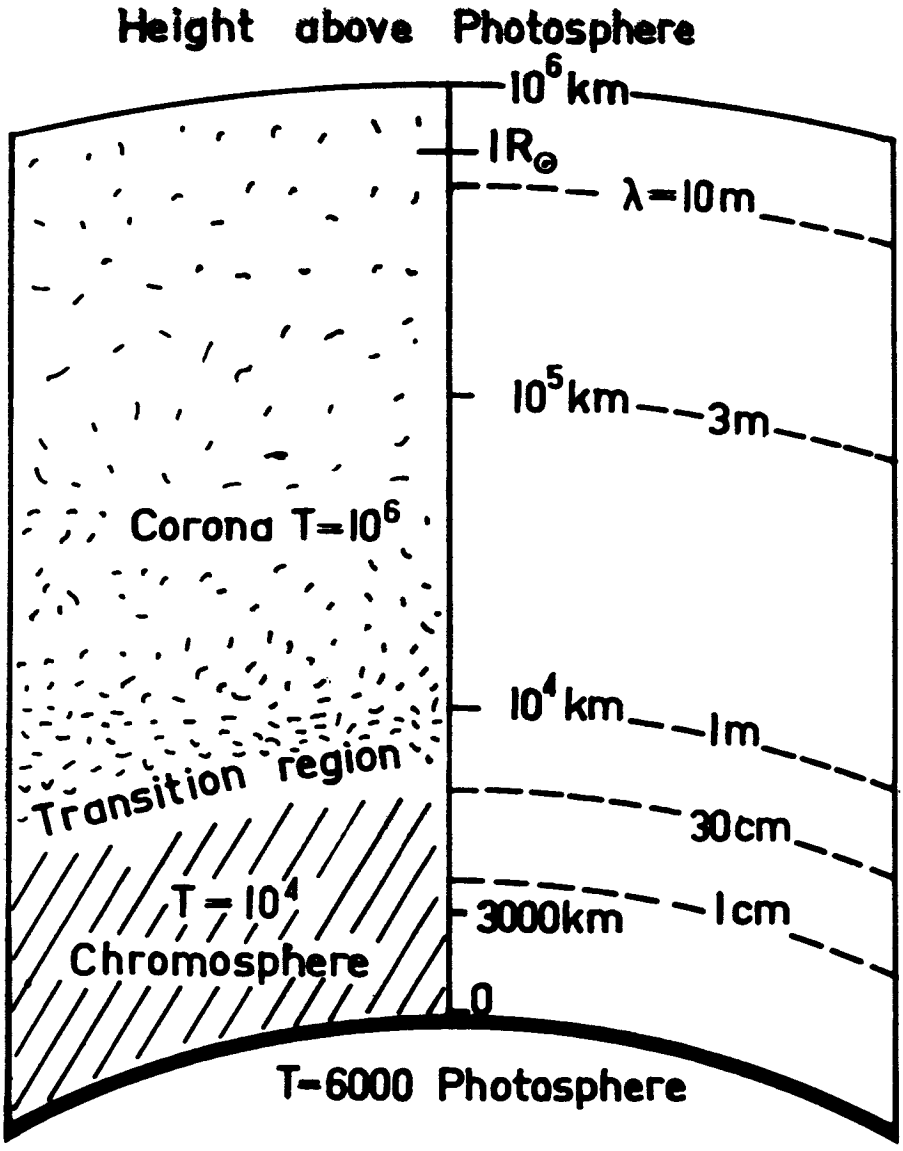


Fig. 1—Approximate heights of transparency limits in the solar atmosphere at various wavelengths. (Height scale distorted)

and flares. Optical observations have yielded a fantastic number of beautifully detailed observations of such phenomena, but physical understanding is surprisingly limited. Radio observations give further information about disturbances in this region, many of them previously unsuspected phenomena, but fortunately, at least in some cases, physical interpretation is keeping pace better.

The two major contributions of radio astronomy to solar physics relate, firstly to the elucidation of the density-temperature structure in the solar atmosphere, which is the factor determining the thermal radio emission from each point on the solar disk, and, secondly, to the complex series of events which accompany solar flares. The first has important terrestrial associations in that the emission of the ultra-violet light and X-rays responsible for ionization in the ionosphere is closely related to the same factors, the second in that flares are associated with the further emission of X-rays, with the ejection from the Sun of the charged particles responsible for our aurora and ionospheric disturbances, and with the emission of solar cosmic rays.

Before discussing these in detail I shall give a brief account of the origins of solar radio astronomy and a description of the general features of solar radio waves and of the theory of their emission. I shall then return to the two special lines mentioned above and conclude with a brief account of information obtained about the outer corona by observations of the scattering of the radio waves from radio stars when these pass close to the Sun.

2. Origins of Solar Astronomy

Soon after Hertz's pioneer work in radio, speculation arose as to the possible existence of solar radio waves. Attempts were made to detect them, but the techniques of the day were hopelessly inadequate. By the 1920's, when adequate techniques first became available, the idea of looking for solar radio waves had been lost. They were actually discovered independently in the year 1942, by Southworth (1945) and Hey (1946) in the United States and England respectively.

It is interesting to note the way in which scientific discoveries may be divided into two classes: (1) the unexpected discovery, made often in the course of observations aimed at another objective, and (2) the verification of an idea derived by a man of wide perspectives from known facts. The two pioneering discoveries of solar radio waves illustrate these two categories. Southworth, from his knowledge of thermal emission processes, believed that the Sun should emit appreciable amounts of radio waves, especially at short wavelengths, and, having available newly developed microwave equipment, he turned this on the Sun and, at 1, 3 and 10 centimetres, he recorded solar emission. He then compared his observed intensities with that expected theoretically from thermal emission from a black body at 6000°K and missed a major further discovery through making a mistake. He incorrectly evaluated the theoretical emission and concluded that, at all three frequencies, his measurements indicated 6000° thermal emission. Actually, his longer-wavelength measured intensities considerably exceeded the 6000° value, and he missed the clue which he himself had

provided to the later development in which measurements of intensities at successively longer wavelengths are used to determine the temperature at successively higher levels in the solar atmosphere. It turns out that the solar atmosphere is highly irregular both in time and space and subsequent developments have rested very much on the outstanding series of daily 10.7 centimetre intensity observations made by Covington and his colleagues at Ottawa, and on the extraordinary development of high angular resolution techniques, culminating in the three-minute-of-arc beam radioheliograph, by Christiansen and his colleagues at Sydney. (Christiansen, Mathewson and Pawsey, 1957)

Hey's discovery falls in the "unexpected discovery" category. In February 1942, English metre-wavelength radar sets experienced what seemed to be severe jamming, suspected to be of enemy origin. Hey investigated the circumstances and found that the effect was due to intense solar emission associated with an exceptionally large sunspot present at that time. Hey's discovery was the starting point for the subsequent series of studies of disturbances in the solar atmosphere which are manifested by bursts of radio emission.

In these studies Wild's and his colleagues' observations at Sydney of the spectra of bursts, including recognition of types in which the frequency band drifts with time (Wild and McCready, 1950), the discovery of second harmonics (Wild, Murray and Rowe, 1953), and the change in position of the source with time and frequency (Wild, Sheridan and Trent, 1959), have played a major part. Boischot and Denisse's (1957) recognition of the type IV burst, and its association with the generation of solar cosmic rays, is also a landmark.

3. General Features of Solar Radio Emission and the Instruments Used for Observation

The feature of solar radio waves which differentiates them most strongly from light is their great variability. Whereas the light from the Sun is constant within observational limits, the radio intensity varies over a range of up to a million to one. Further, the rate of variation can be very fast—gross changes in a few seconds are common. The type of variation encountered at different wavelengths is illuminated in Figure 2. At a wavelength of around one centimetre the radiation is almost constant, excepting for occasional bursts typically of several minutes duration. In the range from about three centimetres to one metre, a new feature appears. In addition to the bursts, which are more intense and more frequent, the level fluctuates in periods of the order of a month over a range of up to about two to one. At metre wavelengths a new regime sets in; the relatively slow variation of general level disappears and there are, at certain times, very large numbers of increases with durations ranging from a second to a few days. Bursts of a few seconds duration are a remarkable feature to be found in the emission from a star.

These various increases are due to a variety of different types of disturbance on the Sun and we clearly require some sort of classification which

TABLE I. RECOGNIZED TYPES OF SOLAR EMISSIONS

No.	Component	Typical Duration	Spectrum	Region of Origin	Degree of ⊙ polarization	Typical T_b (°K)	Emission Mechanism	Remarks
1	Quiet Sun	——	continuum	whole Sun	random	10^4 at 1 cm. to 10^6 at 1 m.	thermal	Related to structure of solar atmosphere
2	Slowly varying component	months	continuum (≈3 cm.—1 m.)	bright regions (radio plages) at base of corona	random at 20 cm. partly ⊙ at 3 — 10 cm.	up to 10^6	thermal	
3	Microwave burst	ten minutes	continuum (≈1 cm.—50 cm.)	small region at base of corona	random or partly ⊙	10^8	Synchrotron ? Bremsstrahlung ?	Associated with flares
4	Type IV burst	hours (very smooth)	continuum (50 cm.—10 m.→)	region high in corona ; often moving	usually random first, then weak ⊙	10^9	Synchrotron	
5	Type II burst (slow drift)	ten minutes	narrow band drifting down in frequency (≈0.2 Mc/s per sec) 2nd harmonic common (1m.→ 10m. →)	region moving from low to high corona (vel.≈1000 km/s)	usually random	10^9	Plasma oscillations	
6	Type III burst (fast drift)	ten seconds	narrow band drifting down in frequency (20Mc/s per sec) 2nd harmonic common (1m.→ 10m. →)	region moving from low to high corona (vel.≈c/2	usually random some ⊙	10^9	Plasma oscillations	
7	Noise storm (continuum and type I or "storm" bursts)	disturbed period, days ; bursts, seconds	continuum plus narrow band bursts (1m.→ 10m. →)	mid-corona	usually strongly ⊙	10^9	Plasma oscillations	Least understood

will be related to their origin. The currently recognized classification into components attributed to physically distinct phenomena is given in Table I. The separation is jointly based on (1) the type of variation with time and (2) spectral features. Table I also includes notes on other observed features and present ideas on mechanisms of origin.

In column 7 intensity is specified in terms of brightness temperature, T_b. This is defined as the temperature of a black body which would give thermal emission equal in intensity to that observed. The values for components 3 - 7 are very rough representative figures; actual values vary enormously from just observable to orders of magnitude higher. A feature of using brightness temperature for this purpose is that the occurrence of values in excess of the temperature of the emitting body immediately indicates a non-thermal emission mechanism.

Circular polarization, either partial or nearly complete, see column 6, is frequently observed in solar radio emission. It is important as being a certain indicator of the existence of magnetic fields, either at the place of origin or along the path to the observer. However, except in the case mentioned in the succeeding section on thermal emission, it has not yet been possible to use radio data to measure magnetic fields on the Sun.

The radiation from the Sun is complex and there is no reason to suppose that other distinct components may not be recognized in the future.

The seven components in this Table fall into three natural divisions: (1) and (2), thermal components characterized by brightness temperatures in the range of temperatures known to exist in the solar atmosphere; (3), (4), (5) and (6), non-thermal components associated with flares; and (7) which is non-thermal but is still little understood.

The recognition of the various components is not always easy. No. 1, the quiet Sun background, is simple in concept and indeed, at times of sunspot minimum it may be the sole component present. However at times of greater solar activity it is often difficult to recognize and reject other components, particularly the slowly-varying one (2). The separation of (1) and (2) can best be done if high directivity data are available showing the bright regions giving rise to (2). Microwave bursts are immediately recognizable on centimetre or decimetre single-frequency records from their short duration, minutes or tens of minutes, because the background due to (1) and (2) is so steady. The metre-wavelength components are very difficult to recognize individually without records of dynamic spectra, i.e. a continuous record of the variation of intensity with both wavelength and time. Given these, a type IV burst is recognized by its wide frequency band, extended duration and absence of short-period bursts; types II and III by their characteristic frequency drifts; and type I bursts by their very narrow frequency band (a few Mc/s). The distinction between the continuum of a noise storm and a type IV burst is still a controversial subject. Typical spectra for these bursts are shown in Figure 3.

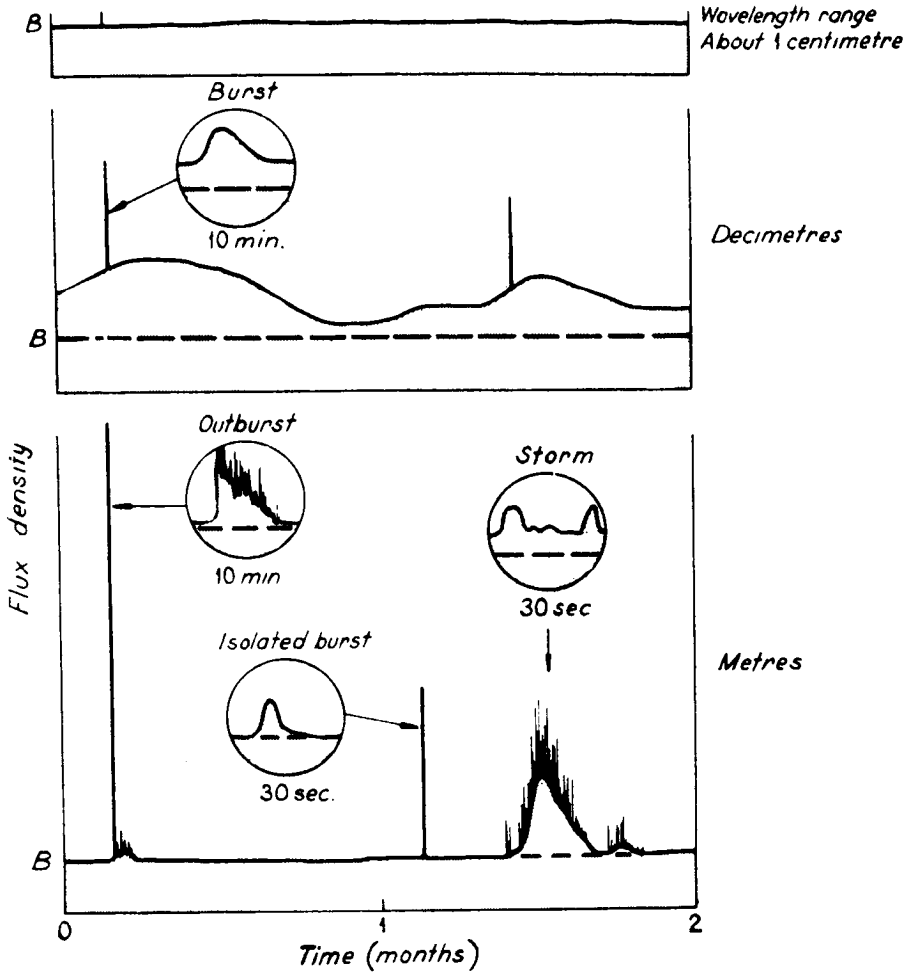


Fig. 2—Idealized records of solar radio waves in representative wavelength ranges. The base level B increases approximately linearly with wavelength from an apparent disk temperature of 10^4 °K at 1 cm. to 10^6 at 1 m. (Pawsey and Bracewell, 1955)

I shall not in these lectures have time to describe the instruments used to record solar radio waves, but I must outline the type of observations they are capable of giving. They fall into four broad classes:—

(1) *Single-frequency recorders.* Each consists of a sensitive and highly stable receiver connected to a directive aerial which is made to follow the Sun. It records the integrated flux density from the whole Sun in a band, usually a few Mc/s wide, at the wavelength of the equipment. (See Fig. 4)

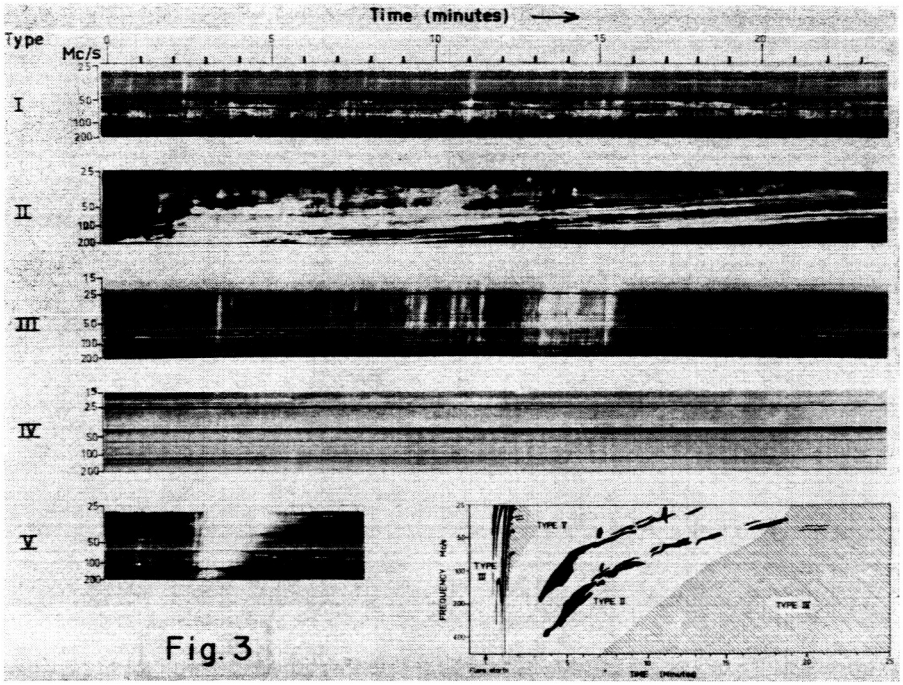


Fig. 3—Samples of recorded spectra of the recognized forms of metre wavelength solar bursts prepared by Wild and McLean for use as standard types. Sequence at bottom right gives sequence observed in a fully developed flare event. (Radiophysics Lab. Sydney)

(2) *Dynamic spectrographs.* Each consists of a broadband aerial, which follows the Sun, connected to a swept-frequency receiver. It is current practice to make each combination cover a restricted range, e.g. two to one, and combine several to get wide coverage. The present Sydney equipment, for example, covers the range 6-240 Mc/s in half a dozen stages. (See Fig. 5 for the aerials, Fig. 3 for spectra.)

(3) *High-resolution instruments.* The only current satisfactory way of delineating a *complex and changing* brightness distribution is to use a pencil beam instrument of adequately small beam-width to scan the Sun. The Christiansen radioheliograph or crossed-grating-interferometer is an example on a wavelength of 21 centimetres. (See Fig. 6) Bracewell's 10-centimetre instrument at Stanford is similar.

For moderately simple distributions, a multi-element interferometer with a fan beam can supply much essential information from one-dimensional scans of the object. Such instruments are numerous (e.g. Sydney, $\lambda = 21$ cm. beam 2′ arc; Nançay, f = 169 Mc/s beam = 3′ arc). (See Fig. 7) Similar results have been attained in Russia with an even higher

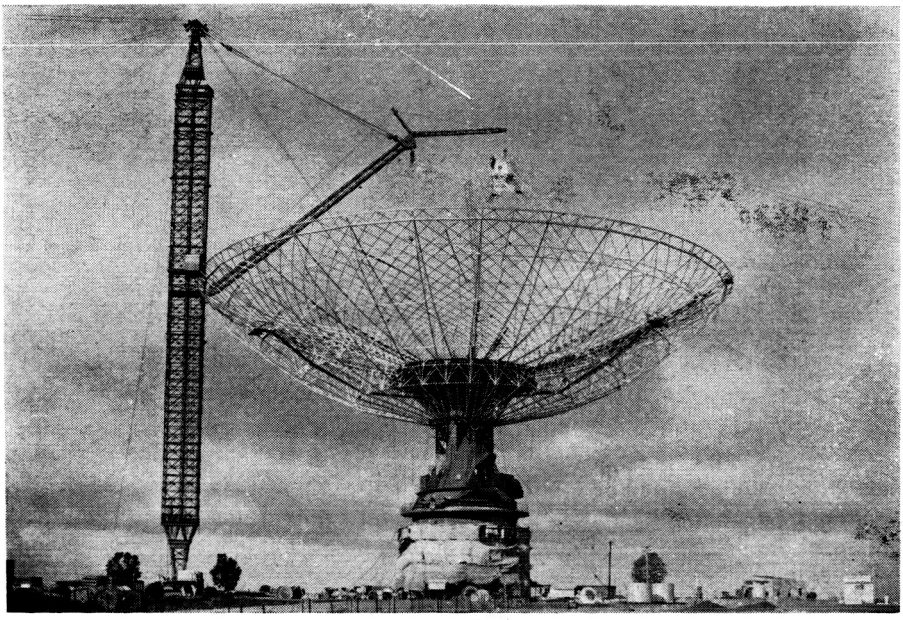


Fig. 4—A recent picture of the 210-foot radio telescope under construction at Parkes, N.S.W. Australia, for the Radiophysics Laboratory.

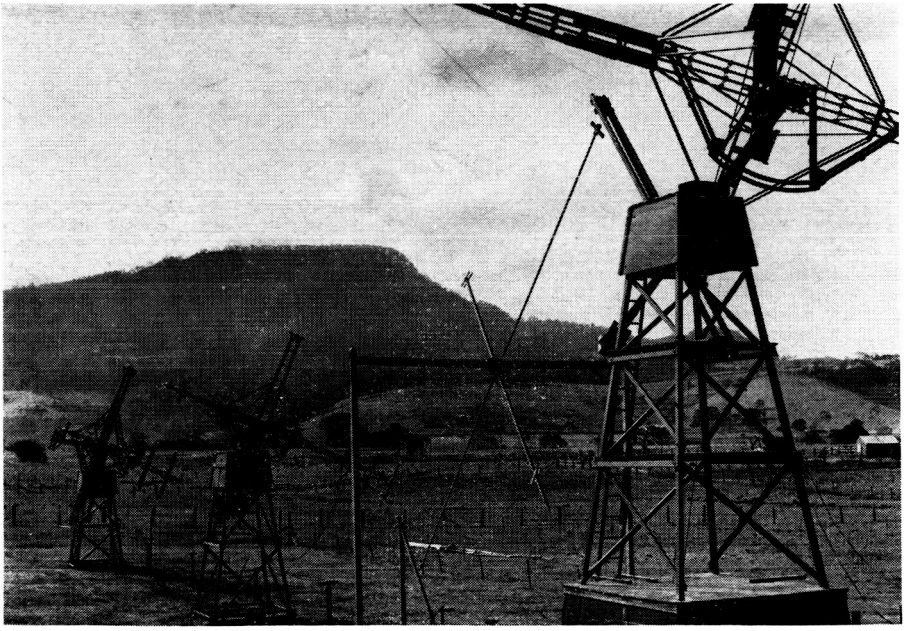


Fig. 5—Rhombic antennas at Dapto, N.S.W. used in the original solar radio dynamic spectrograph. (Radiophysics Laboratory Sydney)

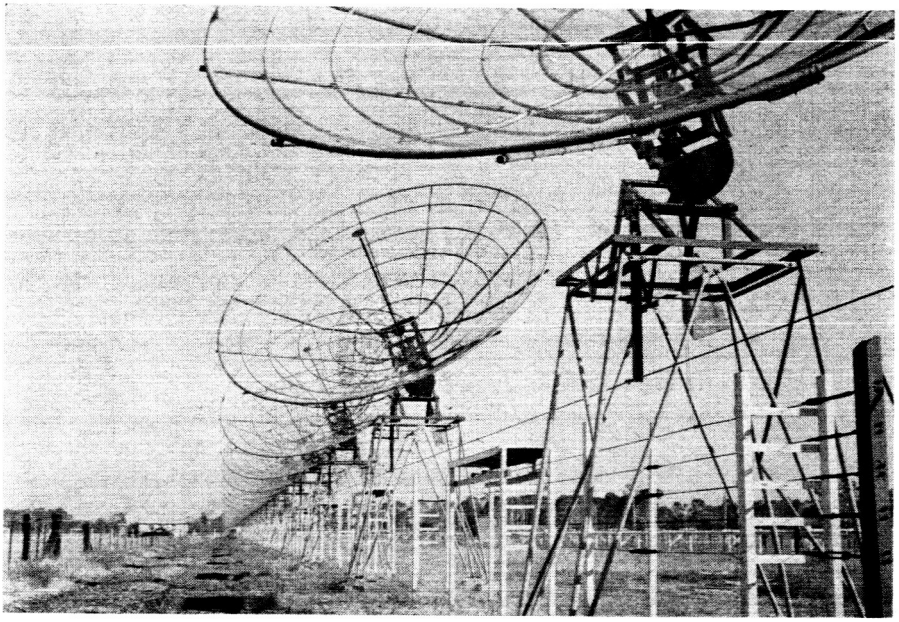


Fig. 6—Some of the 64 19-foot paraboloids used in the original crossed grating interferometer. Length of arms, 1200 ft.; wavelength, 21 cm.; pencil beam width, 3′ arc. (Christiansen, Mathewson and Pawsey, 1957)

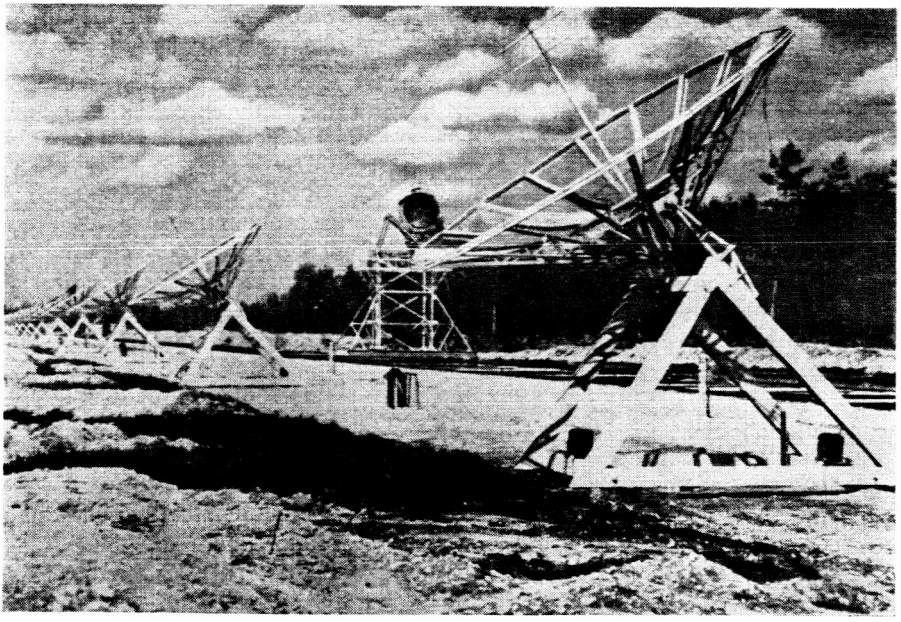


Fig. 7—Some of the 32 paraboloids of "Le Grande Interféromèter de Nançay". Length=1550 m.; f=169 Mc/s; fan beam, 3′ arc. (Blum, Boischoit and Ginat, 1957)

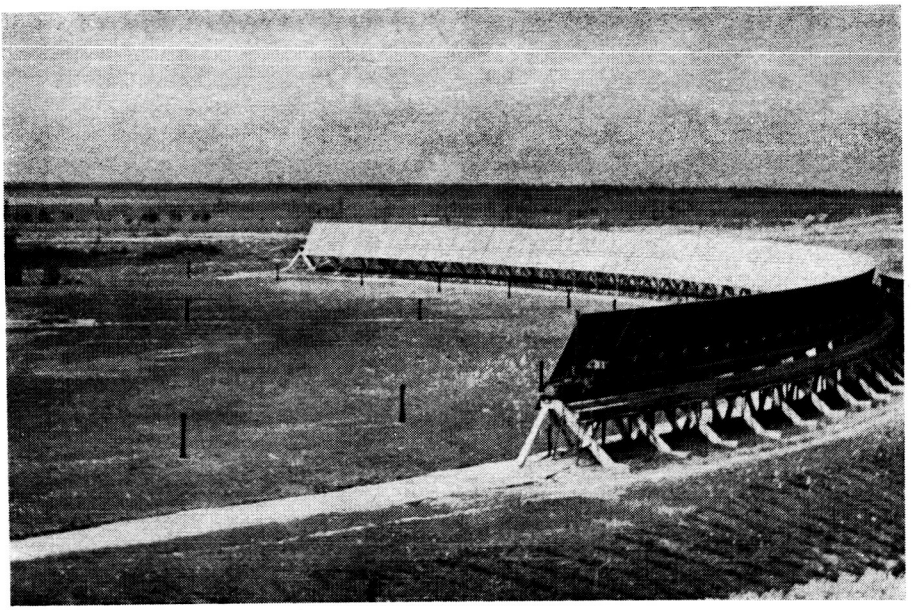


Fig. 8—The Pulkovo radio telescope. Beam 1′ arc × 1° at 3 cm. (Kaikin and Kaidanovski, 1959)

resolution, 1′ arc, with the Pulkovo radio telescope which is built up of 90 reflecting plates arranged in the arc of a circle. (See Fig. 8)

Further, if the distribution of brightness is quite simple, e.g. a single bright blob, a two-element interferometer can supply information on its position and size. Intense solar bursts completely overshadow their background and are effectively studied in this way.

(4) *Polarimeters.* Circular polarization is most commonly observed. Two distinct principles are used. In one case, the relative response on aerials of appropriate polarization is measured. In the other, the correlation is observed, as a function of phase delay, of the signals from two aerials of orthogonal polarization (e.g. linear at right angles).

The above instruments commonly give single-frequency information. Wild's group have broken through this restriction in developing, in conjunction with their swept-frequency receiver which gives spectra, swept-frequency interferometers which give position, size, and polarization of a radio burst as a function of frequency (Wild and Sheridan, 1958).

4. Mechanisms of Emission

It is now clear that there is a relatively steady background of thermal emission above which shorter-duration, non-thermal, contributions appear sporadically. The theory of thermal emission is known and predictions of

its expected intensity, based on the approximately known distribution of electron density and temperature in the solar atmosphere, were made by Martyn (1948) and Ginsburg (1946). On comparing these with observations over the available wavelength range, it was clear that the thermal emission could be recognized, although it was overshadowed at times by the intense non-thermal components. The observations posed two problems: the separation of the thermal and non-thermal components, and the mechanism or mechanisms responsible for the latter.

Thermal emission. The gas of the solar corona and upper chromosphere is predominantly hydrogen and fully ionized—a plasma of electrons and protons. Its temperature ranges from a million degrees in the corona to ten thousand in the mid-chromosphere. In such a gas the predominant thermal emission mechanism is the emission from electrons accelerated at collisions with protons—free-free transitions in the language of the quantum theory. This mechanism is essentially that responsible for "bremsstrahlung". In evaluating the emission, we utilize Kirchoff's reciprocal law of absorption and emission. In an element of a ray in which the increment of optical depth is $\Delta\tau$ (i.e. power attenuation $e^{-\Delta\tau}$) and the temperature T, the emission contributes ΔT_b to the brightness temperature where $\Delta T_b = T\ \Delta\tau$. This is reduced before arrival at the observing point by absorption in the ratio $e^{-\tau}$, so that the actual observed brightness temperature T_b is given by

$$T_b = \int_0^\infty e^{-\tau}\ T\ d\tau,$$

where τ is the optical depth measured from the observer to each point. Now, for ionized hydrogen of temperature T and density N, the absorption coefficient K is given by

$$K = \frac{A\ N^2}{f^2 T^{3/2}}$$

where A is a nearly constant parameter, known for the given conditions, and f the frequency. The emission from a thin slab ($\Delta\tau << 1$) of thickness t is then given by $\Delta T_b = TKt$, i.e.

$$T_b = T\ \frac{A\ N^2}{f^2 T^{3/2}}\ t = \frac{A\ N^2}{f^2\ T^{1/2}}\ t. \qquad (1)$$

The unexpected feature is that the absorption coefficient decreases with increasing T faster than T increases so that increase of temperature decreases, rather than increases, the received intensity.

The above discussion neglects possible magnetic fields and assumes an electron-proton gas in equilibrium. Magnetic fields introduced the two modes of the magneto-ionic theory with their attendant circular or elliptical polarization. One important effect is that a magnetic field increases the absorption, and raises the "visibility limit", of the extraordinary ray to a

region of lower electron density. In a region in which the temperature increases with height the extraordinary ray due to thermal emission will thus emanate from a region of higher temperature than the ordinary, and will tend to be of greater intensity. This leads to partial circular polarization. This effect has been utilized to set an upper limit to a general solar magnetic field on the basis of an observed upper limit to the degree of circular polarization of the radiation from one hemisphere (e.g. north) of the Sun. (e.g. Smerd, 1950) The case in which additional electrons of high energy exist in the plasma is discussed later.

Non-thermal emission. An excellent discussion of non-thermal mechanisms was given by J. F. Denisse at U.R.S.I. 1960. In practice the presumed non-thermal components frequently show brightness temperatures far in excess of temperatures known to exist in the solar atmosphere. The mechanisms of emission must then be powerful ones. Such high intensities can in principle be attained in either of two ways: either (1) by coherent motion of groups of electrons (electric currents) which need not have individual velocities greater than the normal thermal ones, or (2) by emission from incoherent high-energy electrons. Of the three mechanisms proposed (see Table I), "plasma oscillations" falls in category (1), while "synchrotron emission" and "bremsstrahlung" fall in (2).

Denisse remarks that to obtain sufficiently intense radiation from a plasma, there must be a *group of particles having energies far above the mean thermal energies.* These particles can then communicate their energy to the various types of waves which can exist in a plasma. If the coupling is strong, the intensity of the wave concerned will rise to a level appropriate to the effective temperature of the energetic particles; if weak, the particular wave may not be excited. The four waves considered by Denisse include the two electromagnetic waves (the ordinary and the extraordinary ray of the magneto-ionic theory) and two electrostatic waves related to plasma oscillations. The former propagate in a vacuum and can travel in interplanetary space, the latter require the existence of a plasma. In the absence of a magnetic field the former are exactly transverse, the latter longitudinal.

In certain cases the coupling between the particles and one or other of the electromagnetic waves may be strong and also the wave may propagate without serious attenuation to the observer; i.e. the particles can radiate directly. The most clear-cut case is the postulated origin of type IV bursts by the "synchrotron mechanism". The emission occurs when relativistic electrons spiral in magnetic fields. Electrons of low energy spiraling in magnetic fields radiate very feebly except at the gyrofrequency, but waves of this frequency cannot escape because they cannot propagate. However as the energy increases and the electron velocity approaches the velocity of light, the emission increases greatly and extends over a series of harmonics approximating a continuum. Thus pockets of relativistic electrons imbedded in magnetic fields high in the corona, which are the conditions postulated in the type IV case, can radiate freely.

A less clear-cut case concerns microwave bursts. These are observed to originate in regions in the vicinity of the base of the corona. Here propa-

gation conditions should permit escape and the strong coupling could well be due again to the synchrotron mechanism. An alternative is that it might be due to emission at electron-ion collisions as discussed in the case of thermal emission, i.e. it could be bremsstrahlung. The emissivity would be governed by an equation of the form of Equation (1) but with the temperature appropriately increased. The important point is that a simple increase in electron energy decreases the emission, and the only way in which this mechanism could yield such high brightness temperatures would be for the density also to attain very high values. This might happen if the phenomena involved a gross compression as well as a supply of high-energy electrons.

In both these cases, emission should occur over a wide range of frequencies. The synchrotron mechanism involves a magnetic field and the emission is predominantly in the extraordinary mode. In the presence of a magnetic field, the bremsstrahlung mechanism also favours this mode as noted above.

Some of the metre wavelength emissions, e.g. type I, II and II bursts, are narrow-band and for many years there has been speculation as to their being due to plasma oscillations. The difficulty is that, in a uniform plasma, although plasma waves could well be excited by an influx of highly energetic particles or a shock wave, plasma waves cannot propagate beyond the solar atmosphere and there is no first order coupling which would transfer their energy to an electromagnetic wave free to propagate to earth. This difficulty has now been overcome, thanks mainly to the theoretical work of Ginsburg and Zhelezniakov (1957). These workers have shown that, in the vicinity of the critical density region, there can be coupling mechanisms which are effective in the vicinity of the plasma frequency f_0 and its second harmonic $2 f_0$. These couple to the ordinary mode. Denisse points out that, to be effective, these coupling mechanisms require that the two waves concerned should propagate with very low attenuation. This would appear to be the case for metre wavelengths where the coupling region is in the very low density corona; it is probably not so for shorter wavelengths where the coupling region would be in the relatively higher density chromosphere. Thus only metre wavelengths should receive contributions from plasma oscillations and this may well be the reason that fast variations are restricted to metre wavelengths.

These mechanisms imply that for the type IV bursts the various frequencies should come from essentially the same place in the corona, while for the type II and type III bursts the different frequencies should come from a series of levels increasing in height with decrease in frequency. These predictions have been beautifully verified by Wild, Sheridan and Trent (1959) using a swept-frequency interferometer (see Figure 9). Their observations show just this and strongly support the hypothesis.

5. Thermal Emission and the Structure of the Solar Atmosphere

We turn now to the first of our two examples of cases in which radio observations are yielding information on phenomena in the Sun. The theory of thermal emission is known and shows that the height of origin rises from the mid-chromosphere at a wavelength of one centimetre to the high corona

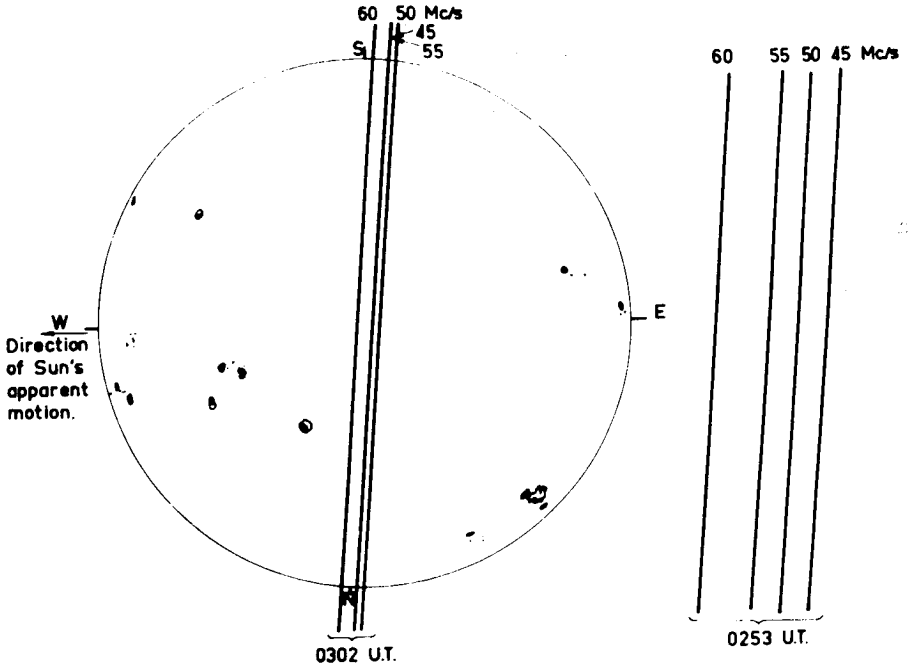


Fig. 9—Observations of the apparent positions of origin of two type III bursts at several frequencies. For the one near the central meridian all frequencies coincide within the experimental uncertainty. The one on the limb shows the lower frequencies originating at progressively greater heights in agreement with the plasma oscillation hypothesis. (Wild, Sheridan and Trent, 1959)

at a few metres. Observations over a range of wavelengths should thus be expected to provide information on the distribution of electron temperature and density in the solar atmosphere.

The first question is the recognition of the thermal contributions, components (1) and (2) of Table I. Figure 10 shows the variation of total emitted radiation over a range of wavelengths from 10.7 centimetres to 1.5 metres, together with the sunspot area, for the period August-November 1947. Occasional short-duration bursts are omitted on the decimetre wavelength curves and only some of the more prominent ones shown by vertical lines on the 1.5 metre one. Taking the 1.5 metre (200 Mc/s) record first, it is clear that there is a well defined base level. This is the level due to thermal emission from the quiet Sun and at metre wavelengths is not difficult to assign. The decimetre wavelength observations, on the other hand, show a continuously varying level, which is actually the sum of that due to the quiet Sun and the slowly-varying component. It is difficult to evaluate the two separately. In the early days attempts were made to separate the two by statistical means, utilizing the high correlation with sunspots illustrated in Figure 10 (Pawsey and Yabsley, 1949), but the best method is to use

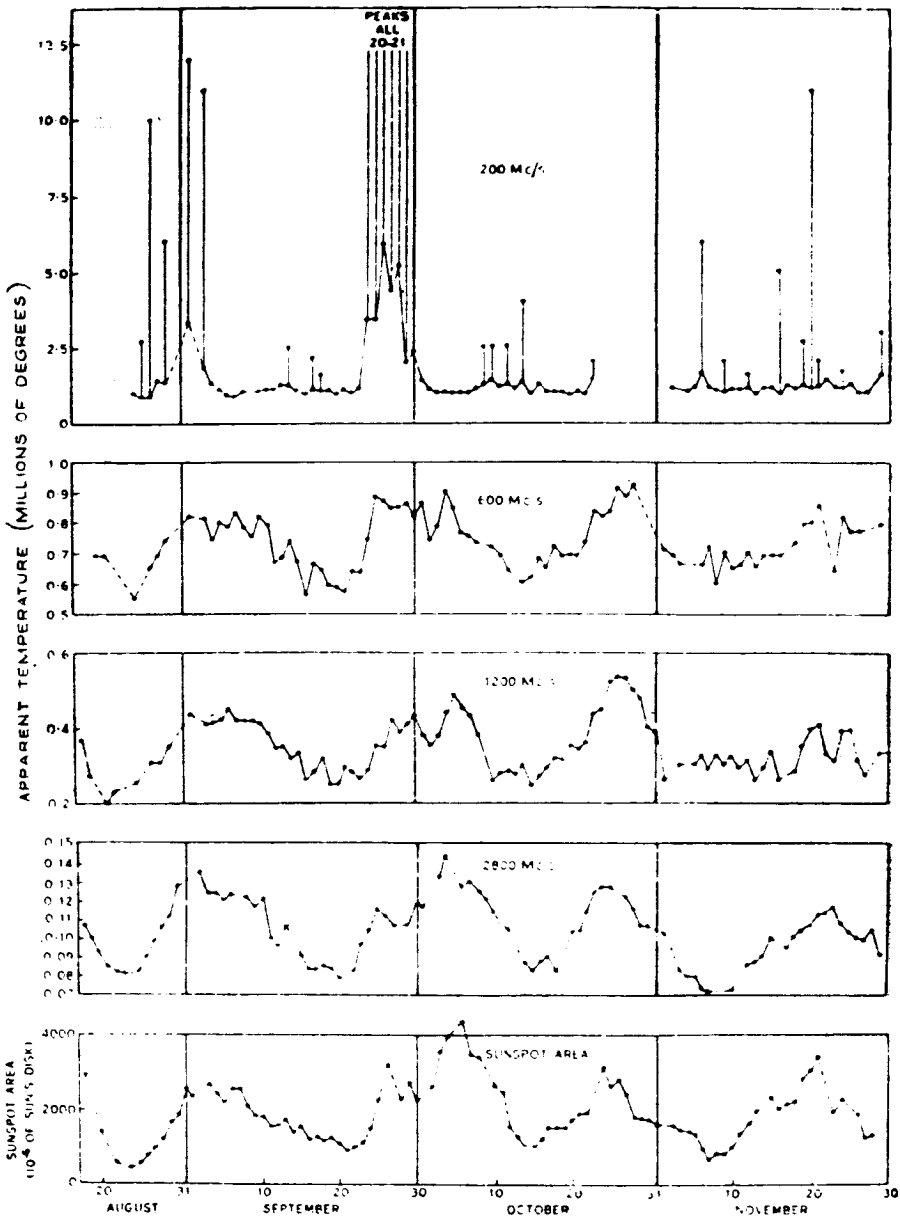


Fig. 10—Observed variation over several months of daily values of the solar apparent disk temperature (equivalent black body temperature of body the size of the visible disk) at frequencies of 200, 600, 1200 and 2400 Mc/s, together with projected sunspot area. Note similarity of variations on the decimetre wavelengths and different character at the metre wavelength. (Pawsey and Bracewell, 1955)

highly directional observations which pick out the individual bright areas responsible for the slowly-varying component (see Figure 11). From a long series of such observations the background can be determined and Figure 12 shows a contour map of brightness over the quiet Sun at a wavelength of 21 centimetres. (Christiansen and Warburton 1956) The interesting features of this map are:—

(1) It extends far outside the visible disk.
(2) It shows banana-shaped *bright* regions at the east and west limb.
(3) It is not circularly symmetric, the north-south section being quite different from the east-west.

These features are readily explicable on the current model, in which the corona is at a temperature of about 10^{6}°K and below its base the temperature falls, reaching about 10^{4}°K in mid-chromosphere. At this wavelength the corona is semi-transparent. A ray coming from the central part of the solar disk thus originates in the chromosphere but receives contributions from both chromosphere and corona. Near the limb in the vicinity of the equator, the slant path through the corona is greater and consequently the corona makes a greater contribution and limb-brightening is produced. Near the poles, on the other hand, although the slant path is again large, the coronal density is less than near the equator and this effect must outweigh the former, since there is limb-darkening. Figure 13 shows a photograph of the corona when exceptionally free from disturbances. The optical brightness pattern shows a striking correspondence with the radio one. In particular the reduced light intensity near the poles is evidence for lower electron densi-

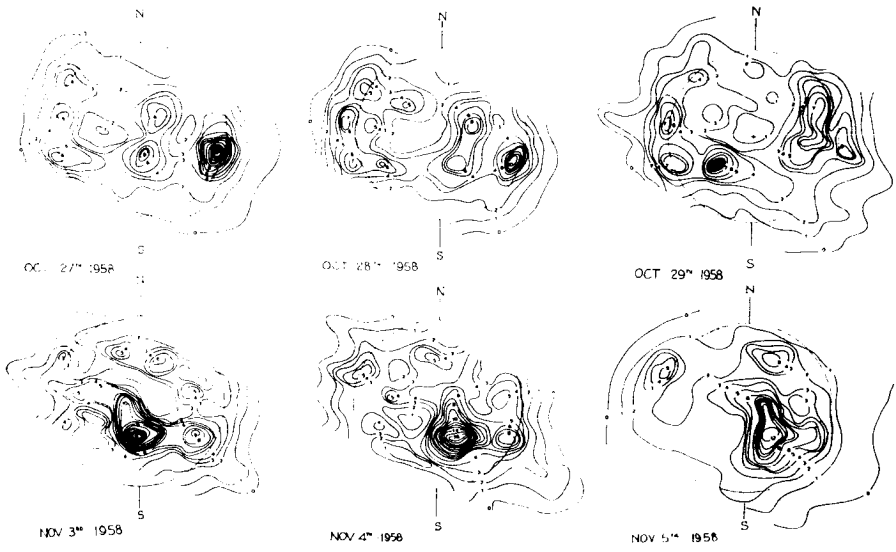


Fig. 11—21-cm. pictures of the Sun (isophotes, unit 10^{5} °K) given by the Christiansen crossed grating interferometer. (Radiophysics Laboratory, Sydney)

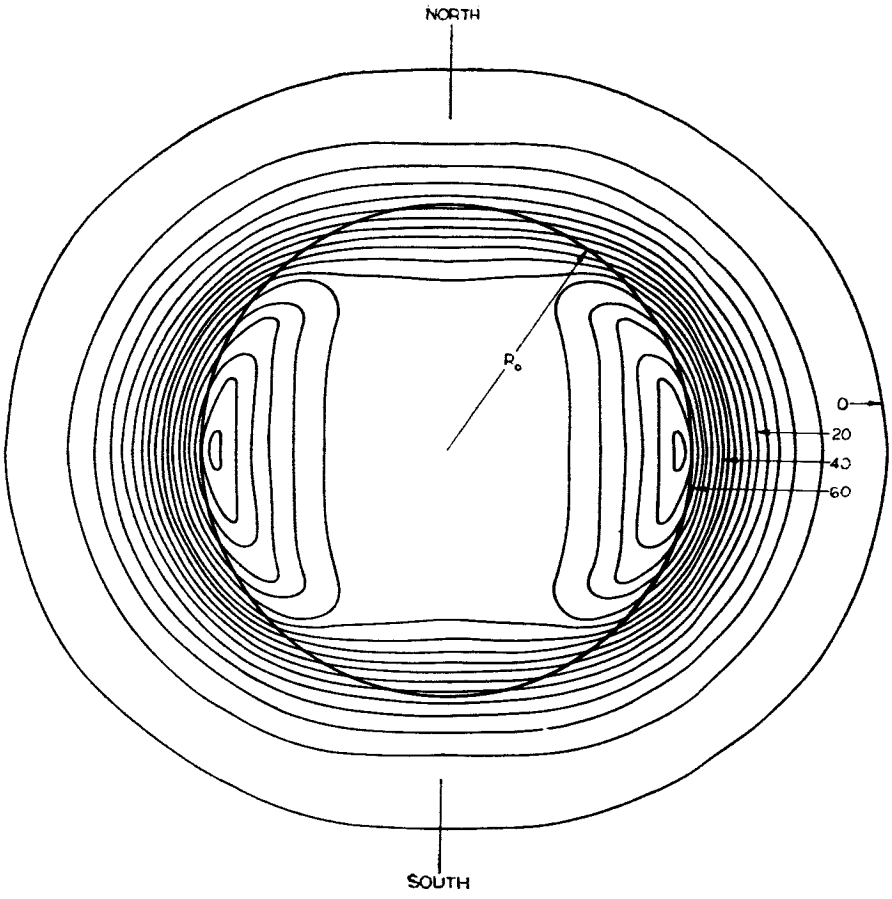


Fig. 12—Brightness contours at 21-cm over the quiet sun, contour intervals 4000°K in T_b. (Christiansen and Warburton, 1955)

ties there. The extension beyond the optical limb is due simply to emission from the corona which extends to great heights. Observations at a single frequency are naturally not enough to derive the temperature-density structure over the whole Sun, although they may be used to check an assigned distribution. Measurements such as those of Figure 12 for a wide range of wavelengths would, however, provide a series of integral equations from which it should be possible to derive this structure. Piddington in 1950 made an attempt on the basis of the limited data then existing. Since that data, observations of brightness distribution have been made covering the range 0.8 centimetres to 7.9 metres and the time may be ripe for a more detailed solution.

Turning now to the slowly-varying component, Figure 14 shows in more detail some of the bright regions in which it originates together with an H α spectroheliogram. I shall call these regions radio plages because of their

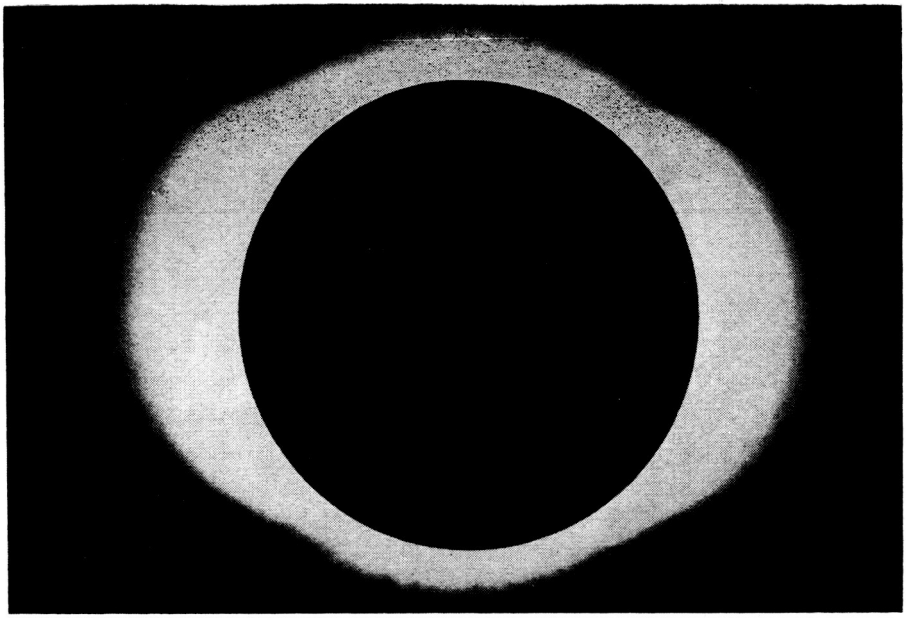


Fig. 13—Photo of the solar corona taken by Waldmeier at the total eclipse of June 30, 1954, a time of exceptional freedom from disturbance. (Christiansen and Warburton, 1955)

similarity and close association with optical (e.g. Ca or Hα) plages. The features of these plages are:—

(1) They are evident in the range 3 centimetres to 1 metre.

(2) They are associated with optical centres of activity and coincide closely with optical plages and with Babcock's magnetic regions. They typically endure for periods of several months—somewhat longer than the associated sunspots.

(3) They move across the disk with solar rotation a little faster than the associated sunspots. This is presumed due to their greater height, and this hypothesis yields heights above the photosphere in the range 20,000 to 100,000 kilometres at 20 centimetres, somewhat less at shorter wavelengths.

(4) They typically have sizes of about 5′ of arc and at 21 centimetres have brightness temperatures which are typically about 5×10^5 and which show a cut-off at between 1 and 1.5×10^6 degrees K.

(5) They show circular polarization in a small area at the shorter wavelengths; none at 20 centimetres.

The specially significant fact here is the limiting brightness temperature of about one million degrees, which is the temperature of the normal corona. All the features agree with the concept that the radio plages are extra-dense

regions in the corona which, because of their excess density, enhance the coronal contribution to thermal emission.

These ideas suggest that, by combining observations of a single plage made at many wavelengths, it should be possible to derive a model giving the structure of the plage. No single observatory makes high-resolution observations over a wide range of wavelengths, but the data is obtainable by combining observations made at various observatories throughout the world. An essay in international collaboration, known colloquially as the "International Geophysical Week" was made along these lines in 1957 and resulted in a paper which must be a near record in diversity of authors:—

Australia—Christiansen, Mathewson, Pawsey, and Smerd
France —Boischot, Denisse, Simon
Japan —Kakinuma
U.S.A. —Dodson-Prince, Firor. (1960)

The paper described a study of a single radio plage, the one in the circle in Figure 14. It contained a great deal of detail but the end point was that it derived the variation of electron density and temperature with height at the centre of the plage. See Figure 15. The problem has been solved in principle.

In this particular case, which we think may be typical, the radio plage consists of an extra-dense region in the lower corona. The temperature distribution shows no departure from normal within the observational uncertainty. Optical observations sometimes show excess temperatures. We do not know what produces this "coronal condensation" or what maintains it fairly stable for periods of a month or so. We suspect it is formed, and constrained, by magnetic fields.

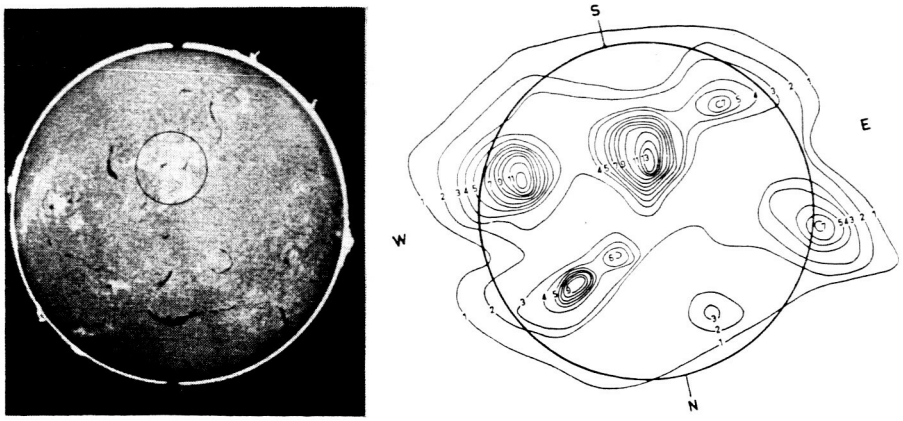


Fig. 14—21-cm isophotes (unit 10^5°K) showing radio plages together with Hα spectrohelipgraph for Dec. 3, 1957. The plage in the circle was the subject of study by an international group. (Christiansen et al. 1960)

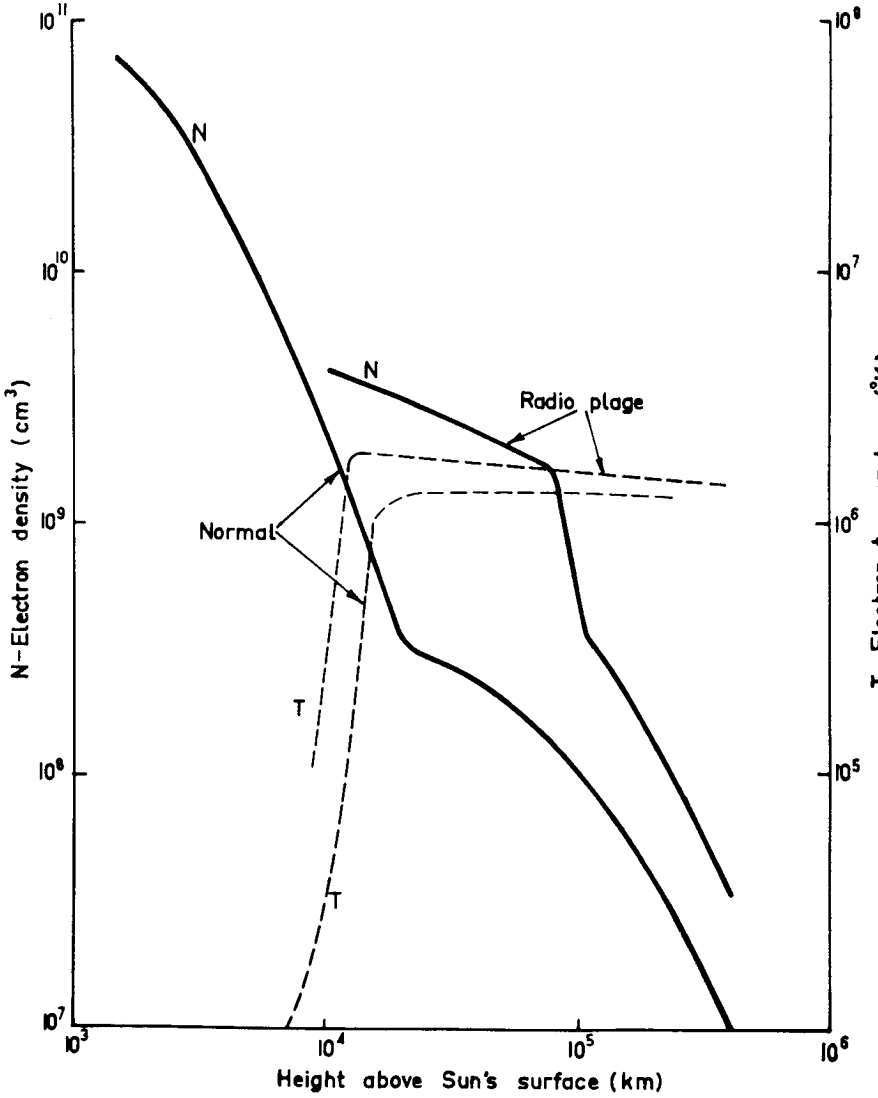


Fig. 15—Values derived directly from radio observations of the variation of electron density and temperature with height in a radio plage together with undisturbed values. (From Christiansen et al. 1960)

These radio plages are of special interest to ionospheric physicists because X-rays which ionize the terrestrial ionosphere are generated in these high-density million-degree regions along with decimetre radio waves. It has been shown, see Figure 16, that the decimetre wavelength (e.g. 10 centimetres) intensity correlates very closely (with a linear relationship) with the ionospheric index $(f_0E)^4/\cos Z$ familiar to ionospheric physicists, as

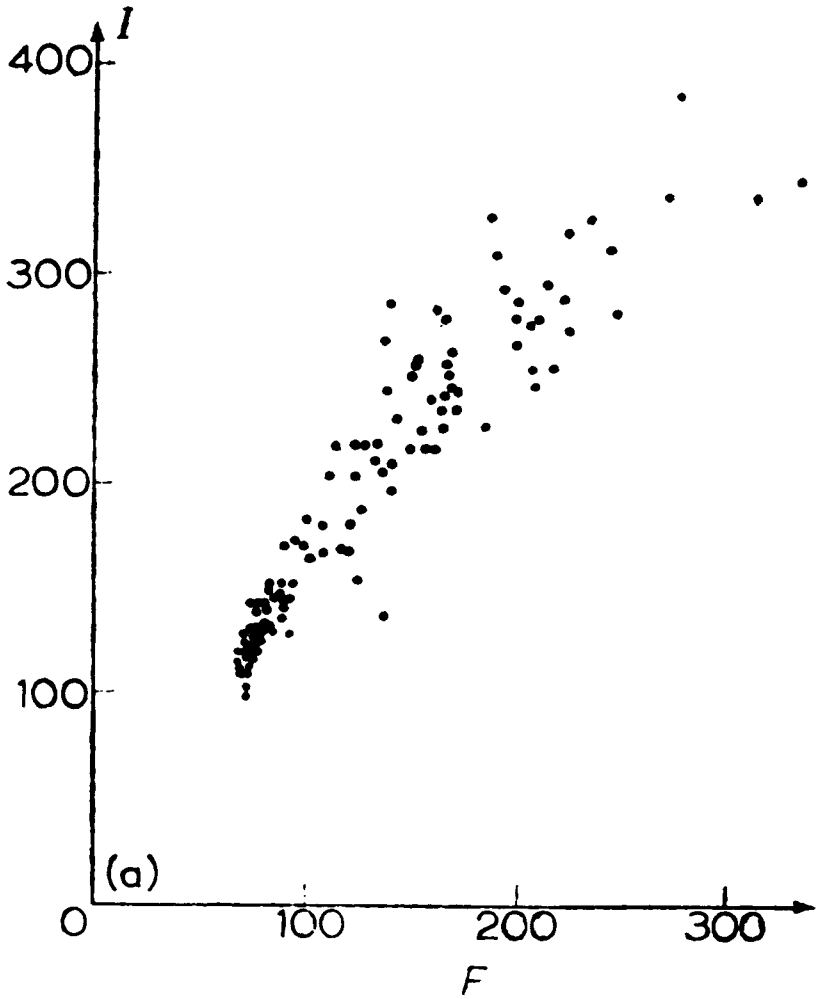


Fig. 16—Scatter diagram, 5-day means, showing the excellent correlation between the 10.7-cm solar flux density, F and the E-layer ionospheric index I given by $I = (f_0E)^4/\cos z$ (Kundu and Denisse, 1958)

a measure of the ionizing radiation incident on the E layer (Kundu and Denisse, 1958). Such a close and simple relationship is a rarity in solar-terrestrial associations.

6. Flares and Their Ramifications

Of the varied forms of solar activity, flares stand out because of their extremely energetic and diverse manifestations. The feature leading to the name, a transient brightening of a small area observed in a spectroheliogram, is a relatively trivial side effect. Something which we cannot see leads to

the emission of intense bursts of ultra-violet light and X-rays, to the explosive ejection from the Sun of vast quantities of gas with velocities of the order of 1000 km/sec, to the generation of high-energy particles (cosmic rays), and to several distinctive forms of radio emission. According to current ideas, the most plausible source of energy for the flare is magnetic. The basic idea is that two regions of magnetic field may collide and mutually destroy one another, the previously existing energy thus being released.

Direct observation of the basic phenomena in flares has proved surprisingly difficult. Optical observations of the Sun show the characteristic monochromatic brightening and frequently the flare is accompanied by the ejection of surge prominences, which do not appear to leave the sun, with velocities up to 1000 or 2000 km/sec. Terrestrial observations show: (1) at the time of the flare, the characteristic S.I.D. due to the ionization of the D-region of the ionosphere by a burst of X-rays (now confirmed by direct observation of the X-rays above the atmosphere); (2) occasional emission of cosmic rays by the Sun, or more frequently, rather less energetic particles (solar protons) which are restricted to the polar regions where they cause excess ionization in the D-region, leading to polar blackouts; and (3) a day or so later, aurorae and magnetic and ionospheric storms due to the arrival at the earth of an ejected cloud of coronal gas. Radio observations are supplying data linking these not obviously related phenomena.

Typical spectra of the metre wavelength types of bursts were shown in Figure 3 and a sequence characteristic of a fully developed flare is also sketched in the lower right-hand corner. The flare is also normally accompanied by a microwave burst without much structure, starting near the beginning of the flare and lasting for ten or twenty minutes. The type IV burst lasts several hours usually. Only occasional flares are accompanied by all these phenomena, but the sequence given is that in which they tend to occur if present. I shall discuss each burst in turn, giving the current interpretation but not attempting to justify the latter.

Type III burst. The cluster of type III bursts emitted at the beginning is evidence for the emission from the flare of a series of disturbances which travel upwards through the corona with velocities of about half the velocity of light to regions of progressively decreasing density, exciting plasma oscillations of progressively decreasing frequency at each level on the way. Wild's suggestion as to the nature of one of these disturbances is that it is a cloud of high-energy electrons ejected upwards.

Microwave burst. This we take to give evidence, at least in the more intense cases, of relativistic electrons in a localized region in the vicinity of the base of the corona. This region may well be the source of the X-rays which cause S.I.D.'s, and of the electrons responsible for the type III bursts. It is probably the heart of the flare, visible in radio waves but probably not in light.

Type II burst. This burst with its relatively slowly diminishing frequency band, again, as in the type III case, evidences a disturbances travelling

upwards. In this case the velocity is about 1000 km/sec and it probably constitutes a vast gaseous mass ejected by the flare. This is presumably the gas which is projected outwards to give terrestrial aurorae etc. a day or so later.

Type IV burst. This long-enduring burst implies the creation, in the high solar atmosphere, of a cloud of high-energy electrons moving in magnetic fields and presumably retained by these fields as are the high-energy particles in the van Allen belts. How it is created is still a mystery, but it appears that it must be a product of the mass of gas of the type II burst in its rush outwards through the corona. It seems that in the type IV burst, and possibly the microwave burst, we see for the first time the generation of cosmic rays.

Now looking at the whole ensemble there are several most interesting features. Firstly, the two moving disturbances, the causes of the type III and type II bursts, are each seen first when they are in the lower corona and the faster of the two appears first. If we assume that each of the two came from a lower level than that at which it was first seen then the observations are consistent with a simultaneous origin at a common level just below the base of the corona. Within the observational uncertainty this is also the level of origin of the microwave burst which originates at about the same time. Are the three, diverse manifestations of a common huge release of energy? Is the microwave burst due to direct emission from high energy electrons formed, and perhaps trapped, at the lower level? Are the type III bursts due to pockets of electrons escaping from the region? And is the type II burst, with its massive ejection of matter, the main explosion wave due to a tremendous release of energy?

Secondly we return to the energy source. In discussing possible emission mechanisms for the microwave burst I mentioned the possibility that the emission might be bremsstrahlung but that, if so, there must be a gross compression in order that the electrons might emit efficiently. Such a compression might well be one of the effects of the release of magnetic energy. If so, there would be a population of very high energy electrons, and probably other particles, in a plasma of reasonably high density. These are the conditions which favour the initiation of a thermo-nuclear reaction. Are flares then, manifestations of solar hydrogen bombs?

7. Scattering of the Radiation From Radio Stars and the Over-all Extent of the Corona

In this last section I shall describe briefly a solar investigation in which the radiation does not come from the Sun itself but the observed features are illuminated by the radio waves from a radio star. In the early 1950's it was shown by Hewish in Cambridge (Hewish, 1959) and Vitkevitch in Russia (Vitkevitch, 1959), working independently, that the radio star Taurus A, the Crab Nebula, disappeared as it passed close to the Sun, which it does each June (owing to the annual motion of the Sun). The effect is due to

scattering of the radio waves by inhomogeneities in the corona and consists of a spreading and dimming of the radiation, closely analagous to the optical effects observed when we view a street light through a fog. The observations have been used to gain information on the shape of the irregularities and of their variation in space and time during the sunspot cycle, but I shall restrict myself to a single series of observations made by O. B. Slee at Sydney (unpublished) with the object of delineating the overall extent of the observable corona. He observed a relatively large number of radio stars, with paths relative to the Sun shown in Figure 17, with an interferometer of very wide spacing (10 km) working on a wavelength of 3.5 metres. The response on this instrument decreases rapidly with any spreading of the source over about 1′ arc. The sources disappeared owing to scattering as they approached the Sun, and Figure 17 shows the over-all picture of visibility obtained in this way. It will be seen that the effects of the corona can be seen out to about 100 solar radii. This is much further than it can be seen optically (about 10 or 20 solar radii). The radio observations clearly show that coronal irregularities, and hence the corona itself, extend out to this great distance but they do not yet give us a precise picture of the nature of the corona in such regions. The degree of scattering is governed jointly by the excess electron density in the irregularities, and their size and shape. We require information on one of these factors from an independent source before being able to utilize fully these most intriguing observations.

8. Conclusion

Solar radio astronomy today is nineteen years old. What are its contributions to science? I should list the main ones as follows:—

(1) It has shown us remarkable new phenomena associated with flares and promises a much better understanding of their mechanisms, including their ejection into interplanetary space of the characteristic charged particles: solar cosmic rays and "solar protons", and the particles responsible for the aurora, magnetic storms etc.

(2) It has provided us with a very powerful new method for determining the structure of the solar atmosphere, and this promises a much better understanding of the emission by the Sun of X-rays and far ultra-violet light.

(3) It has made remarkable contributions to radio technology in the high resolution field. Instruments developed for solar work now exist with pencil beams of 3 minutes of arc and one with a fan beam of 1 minute.

(4) It has provided a very strong stimulant to optical solar astronomy, from which it has likewise benefitted.

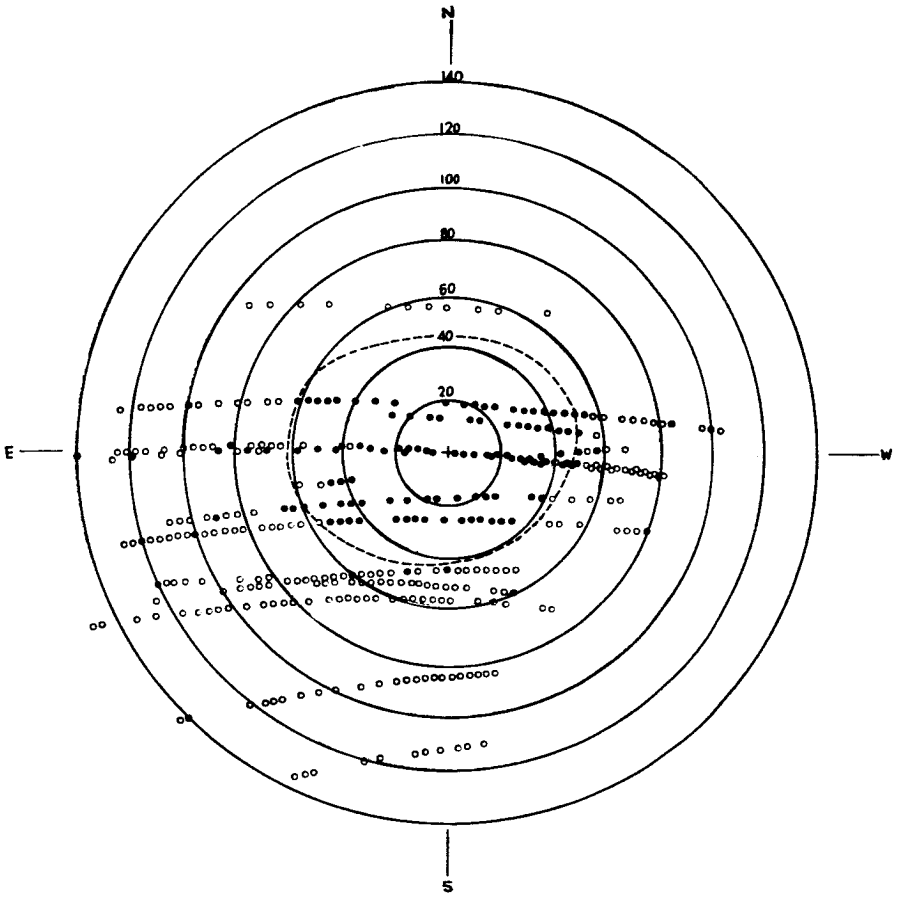


Fig. 17—Scattering of the 3.5 m radiation from various "radio stars" as they pass close to the Sun (at centre); filled in circles indicate significant scattering; open circles, none. Note consistent effects within the dashed ellipse; sporadic effects to 100 R_0 (O. B. Slee. Aust. J. Phys. in press)

References

Blum, E. J., Boischoit, A. and Ginat, M. (1957) Ann. d' Astrophys. *20*, 155.
Boischoit, A. and Denisse, J. F. (1957) C. R. *245*, 2194.
Christiansen, W. N., Mathewson, D. S. and Pawsey, J. L. (1957) Nature *180*, 944.
Christiansen, W. N. and Warburton, J. A. (1955) Aust. J. of Phys. *8*, 474.
Christiansen, W. N. et al. (1960) Ann. d' Astrophys. *23*, 75.
Denisse, J. F. (1960) Lecture given to Com. V at U.R.S.I. General Assembly, London.
Eshleman, V. R., Barthle, R. G. and Gallagher, P. B. (1960) Science *131*, 329.
Ginsburg, V. L. (1946) Comp. Rend. (Doklady) de l' Academic des Sciences de l' U.R.S.S. *52*, 487.
Ginsburg, V. L. and Zheleziakov, V. V. (1959) Paris Symposium on Radio Astronomy, (I.A.U. No. 9) Stanford Univ. Press p. 574.
Hewish, A. (1959) Paris Symposium on Radio Astronomy, (I.A.U. No. 9) Stanford Univ. Press p. 268.
Hey, J. S. (1946) Nature *157*, 57.
Kerr, F. J. (1952) P.I.R.E. *40*, 660.
Khaikin, S. E. and Kaidanovskii, N. L. (1959) Paris Symposium on Radio Astronomy, (I.A.U. No. 9) Stanford Univ. Press p. 166.
Kundu, M. R. and Denisse, J. F. (1958) J.A.T.P. *13*, 176.
Martyn, D. F. (1948) P.R.S.A. *193*, 44.
Pawsey, J. L. and Bracewell, R. N. (1955) Radio Astronomy, Oxford, Claredon Press.
Pawsey, J. L. and Yabsley, D. E. (1949) Aust. J. Sc. Res. A. *2*, 198.
Piddington, J. H. (1950) P.R.S.A. *203*, 417.
Smerd, S. F. (1950) Aust. J. Sc. Res. A. *3*, 265.
Southworth, G. C. (1945) J. Frank. Inst. *239*, 285.
Vitkevich, V. V. (1959) Paris Symposium on Radio Astronomy (I.A.U. No. 9) Stanford Univ. Press p. 275.
Wild, J. P. and McCready, L. L. (1950) Aust. J. Sc. Res. A. *3*, 387.
Wild, J. P., Murray, J. D. and Rowe, W. C. (1953) Nature *172*, 533.
Wild, J. P. and Sheridan, K. V. (1958) P.I.R.E. *46*, 160.
Wild, J. P. and Sheridan, K. V. and Trent, G. H. (1959) Paris Symposium on Radio Astronomy (I.A.U. No. 9) Stanford Univ. Press p. 176.

N63-10223

ENERGETIC PARTICLE PRODUCTION BY SOLAR FLARES

KINSEY A. ANDERSON
Department of Physics
University of California, Berkeley

Introduction

At the present time no theoretical basis exists for describing the acceleration of energetic solar particles, their propagation through the solar atmosphere and their storage in the corona or elsewhere. The reason is that the chromospheric flare process is complicated in detail and the condition of the corona is largely unknown particularly above the active region. Therefore at the present time it is necessary to be content with a descriptive approach to the phenomena of high energy particle production by the sun based on observed behavior of the cosmic rays near the earth and from detailed study of the accompanying solor phenomena.

As may be anticipated from this brief preface the first knowledge that the sun accelerates particles to high energy (substantially above thermal stellar temperatures) came experimentally from an apparatus designed to monitor galactic cosmic rays at ground level. The first figure shows this observation made by Forbush with ionization chambers.

There are several important points:

1. The definite flare association. The first flare was seen by optical means while the second was inferred from ionospheric disturbances.
2. The duration of the particle emission is somewhat longer than the duration of the optical flare.
3. The cosmic rays* can arrive at the earth from greatly different positions of the flare on the solar surface. On 28 February the flare appeared near central meridian. On 7 March this active region evidently yielded another large flare. At this time it must have been near or on the west limb.

*The emission of solar particles over a wide range of energies makes it advisable to define fairly precisely the term cosmic rays. Cosmic radiation is characterized by the following: a. The particle energies are many times greater than thermal energies. (For example those associated with the hottest stellar interiors.) b. The chemical composition is required to include species of a considerable range in atomic number, roughly approximating the Universal Abundances. c. The particles ordinarily do not affect the motion of other cosmic ray particles. That is, they are non-interacting due to low density and they do not ordinarily exert collective effects. Possible violation of this latter point might be anticipated in the acceleration region if it is small and in weak interplanetary magnetic fields which would contain a single particle but might be pushed away by a large flux of particles.

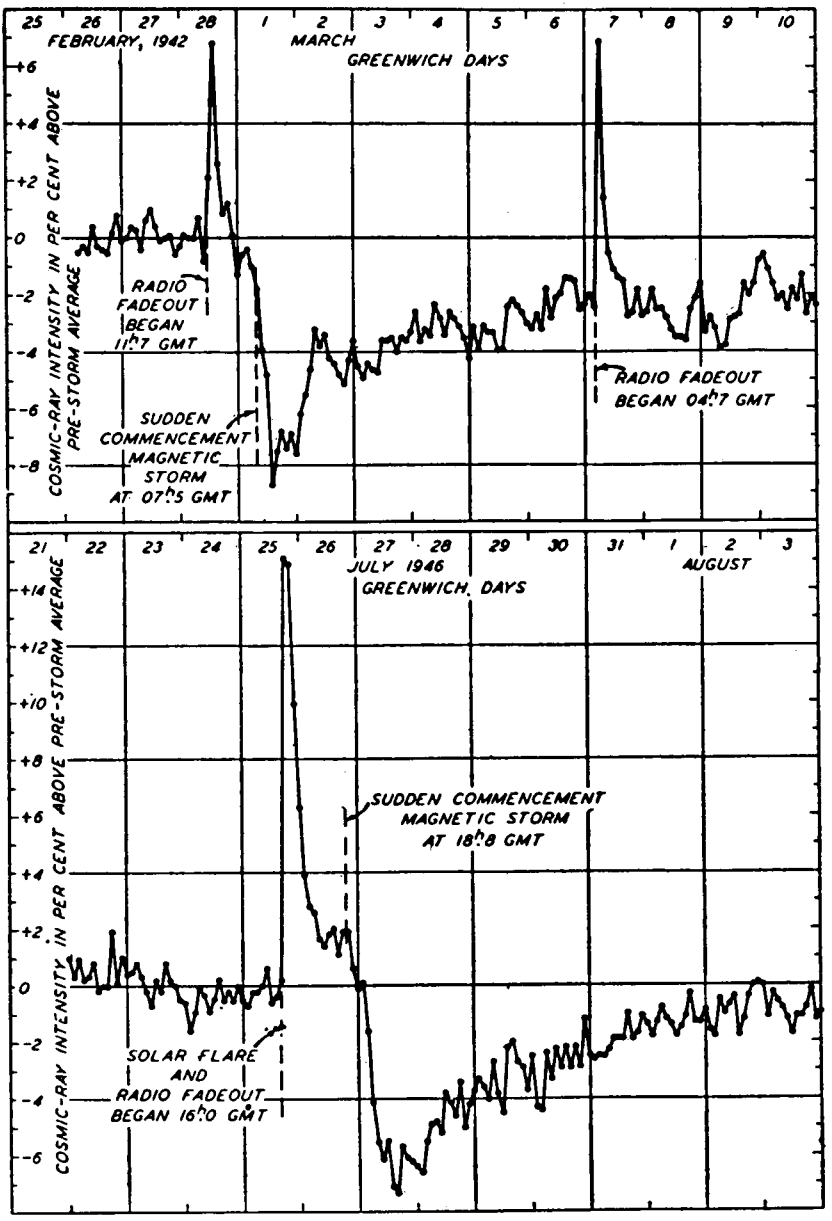


Fig. 1—Three unusual increases in cosmic-ray intensity at Cheltenham, Maryland, during solar flares and radio fadeouts.

On 23 February 1956 a very large ground level event occurred and was thoroughly studied by a then newly introduced device, the neutron monitor, which is far more sensitive to the lower energy particles than the sea level ionization chamber.

One of the most important discoveries during the IGY was that most of the sun's production of cosmic ray particles goes undetected by such devices as the ion chamber and neutron monitor. By means of radio physics techniques and particle counters carried in balloons and satellites it was found that a substantial fraction of optically large solar flares generate high energy radiation which we can call cosmic radiation. Figure 2 illustrates an observation of low energy cosmic rays made at Ft. Churchill on 22 August 1958. Three separate detectors were carried to an altitude of 105,000 feet

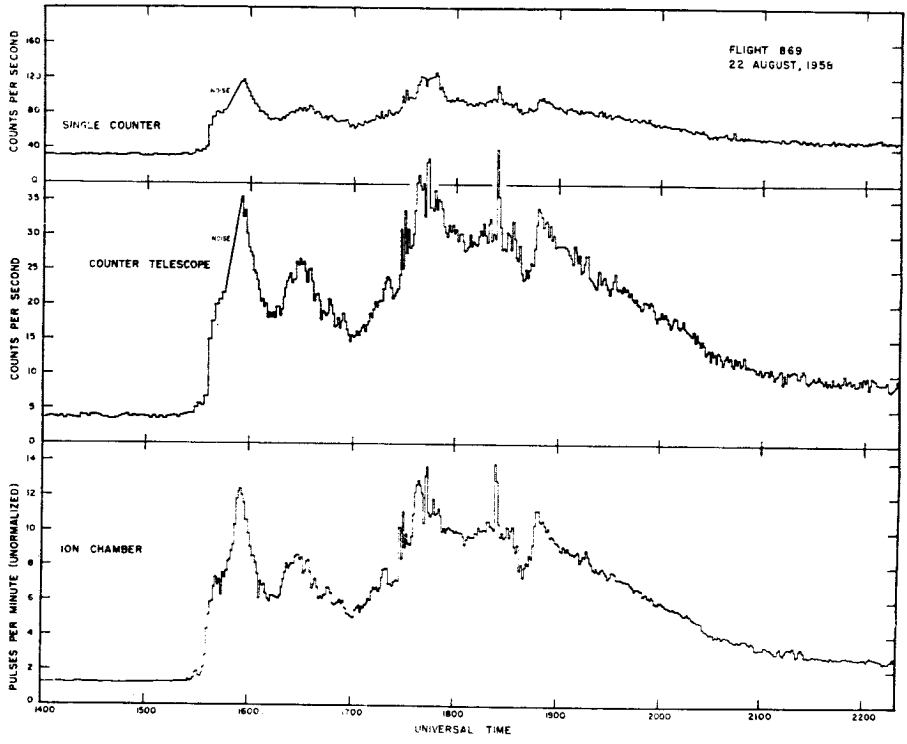


Fig. 2—The solar proton flux following a large solar flare on August 22, 1958 as measured by balloon borne detectors above Fort Churchill, Canada (situated just inside the auroral zone). On this occasion flux of protons above 100 Mev in energy was 20 or 30 times the total galactic cosmic ray flux during the highest peaks.

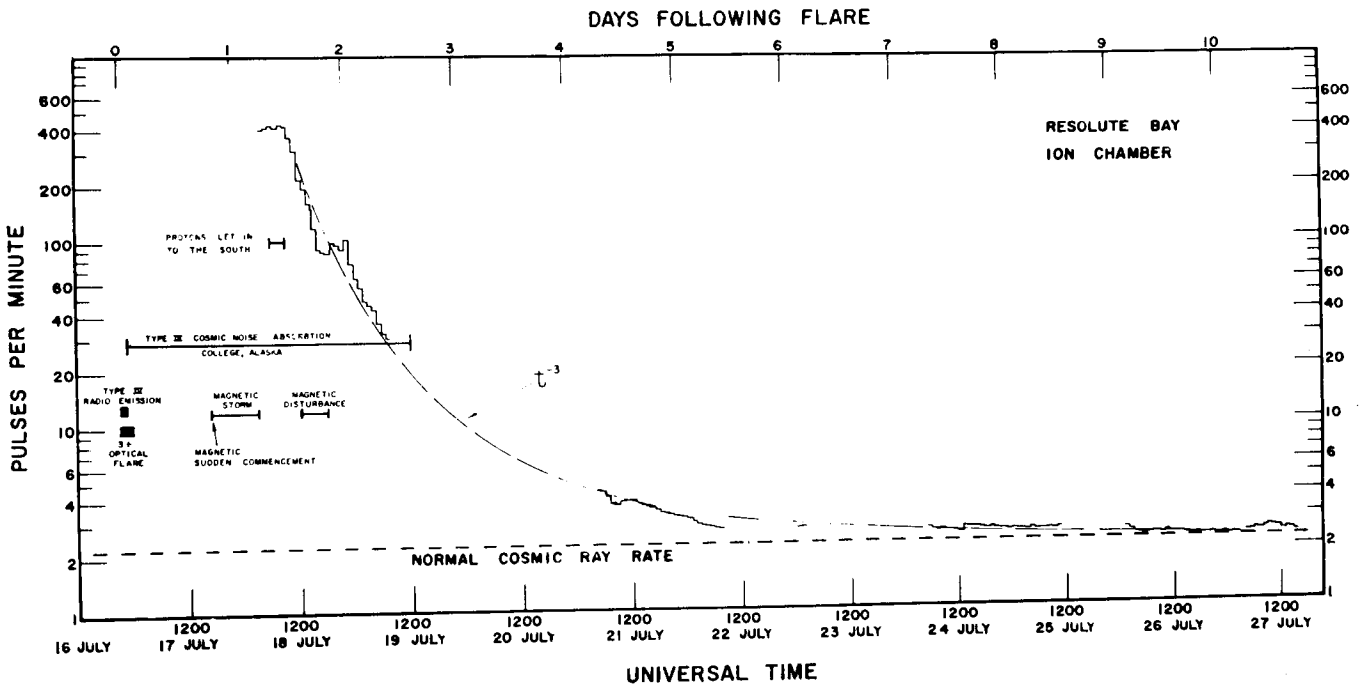


Fig. 3—The third and last of the great July 1959 solar proton events. The solar-particle ionization a day after the flare was at least 200 times the normal galactic cosmic ray ionization at an atmospheric depth of 8 gm/cm^2. The flux was still present after 11 days but quite small for the last 4 or 5 days.

by a plastic research balloon. Some of the notable features on this occasion were the following:

a. The definite flare association with the particles arriving shortly after the flare peak.
b. Large and rapid fluctuations.
c. The nearly smooth tailing off of the particle flux after 1900 UT and continuing past sunset.
d. The particles measured at the balloon consisted mainly of protons of 100-300 Mev energy above the atmosphere.
e. No ground level effects occurred in neutron monitors.
f. There were no marked magnetometer disturbances such as bays nor were there any other auroral manifestations.

Therefore the solar cosmic ray events have no direct connection to the auroral phenomena.

This event led to the rapid understanding of this phenomenon and to ways of more effectively studying them. During 1959 there were rather few solar cosmic ray emissions but these turned out to have enormously high intensity. Figure 3 shows some of the features of one of these events.

It is of particular interest that the active region in which the particle producing flare appeared was at central meridian on the back side of the sun when particles were still being detected at the earth.

Frequency of the Cosmic Ray Emissions

There have been approximately 50 events from 1957 to 1961. This is an average of one per month, but there is a significant bunching, as the same solar active region may produce several events during one passage. About 20% of 2+, 3, and 3+ flares during the IGY produced solar cosmic rays which were received at the earth.

Charge and Energy of the Solar Cosmic Rays Spectrum

The identification of the chemical elements contained in the solar cosmic rays has been achieved by use of photographic emulsions by the University of Minnesota and Goddard Space Flight Center groups. The intensity of the particles and their energy have also been estimated by use of this technique. By means of ionization chambers and other counters, measurements of the intensity and spectrum have also been made. In the fairly large number of events now studied the spectrum is characteristically a steep one and shows a marked flattening below some energy, often about 30 Mev. In one case studied by rockets there were no protons below a few Mev. At the higher energies the spectrum can fairly well be represented by

$$n(E)dE = KE^{-\gamma dE} \quad 50 \text{ to } 300 \text{ Mev}$$
$$\gamma \sim 5$$

In connection with the spectrum studies there is at least one case in which the spectrum must have been much more steep. On 29 August 1957

a balloon at high altitude did not detect solar cosmic rays even though this was the occasion of strong cosmic noise absorption. It is concluded that on this occasion N(>100 Mev), the integral flux of protons above 100 Mev, must have been very small and the flux less than 100 Mev quite substantial.

Solar Regime

Some of the associations exhibited by the solar cosmic ray events are as follows:

a. Nearly all the solar cosmic ray emissions have been associated with large optical flares on the visible disk of the sun. Small flares apparently give undetectable fluxes of cosmic rays.
b. The particle producing flares are quite typical large flares in their optical features.
c. The radio noise emission of cosmic ray producing flares is generally copious but not always exceptional. It is usual for the flare to produce marked type IV continuum radio noise emission.
d. As do nearly all large optical flares, the cosmic ray producing flares generally occur in sun spot groups which are large and complex in visible features as well as magnetically.

Significance of Solar Cosmic Rays to the General Cosmic Ray Problem

At the present time the origin of the high energy galactic cosmic radiation is unknown. Therefore, the possibility should be considered that the flare production of low energy particles provides that raw material for acceleration to much higher energies by "statistical" acceleration processes in the galaxy of the type first suggested by Fermi. The assumption that all stars in the galaxy emit as does the sun is entirely a gross one and, in fact, is probably not even the best that can be done with present day astronomical evidence. However, it is one means of getting a first assessment of the possibility that the galaxy contains copious fluxes of particles in the 10 Mev to 300 Mev region. Particles of this energy might have some difficulty in reaching the earth because of magnetized solar material impeding their propagation yet must be carefully looked for in our environment especially near solar minimum.

INFLUENCE OF SOLAR PHENOMENA ON THE EARTH'S ATMOSPHERE AND MAGNETIC FIELD

SYDNEY CHAPMAN
High Altitude Observatory
Boulder, Colorado

Abstract

Previous speakers have outlined the intrinsic changes on the sun, as shown most visibly by sunspots, prominences, the corona, and solar flares. Their influence upon the earth depends both on the earth's heliocentric direction and its distance from the sun, and also on the sun's geocentric direction.

The fluctuating short-wave radiation of the sun affects the earth's atmosphere mainly over the sunlit hemisphere, where it generates the non-polar ionosphere, and causes thermal motions of the atmosphere. These in turn, along with tidal motions, generate the electric currents whose magnetic field is responsible for the daily magnetic variations observed at quiet times at the earth's surface.

The varying material emission, of solar gas, affects the outer part of the Van Allen radiation belt; and some of it reaches the earth, probably more on the night than the day side. It enters the atmosphere mainly in auroral light. Also it generates powerful electric currents along the auroral zones, of order up to one or two million amperes. They complete their circuit in the atmosphere over the polar caps within the auroral zones, and to a less extent over the great interzonal belt of our globe. They produce the irregular magnetic variations observed at the earth's surface during magnetic storms. The enhancement of the Van Allen belt is associated with the growth of a mainly westward electric "ring" current around the earth. This probably generates neutral lines in the magnetic field above the earth, which seems likely to be closely associated with many detailed features of auroral morphology.

PARTICLES, ELECTRO-MAGNETIC RADIATION, AND DUST

SIEGFRIED F. SINGER
Associate Professor of Physics, University of Maryland
College Park, Maryland

Abstract

A gas of low density may be treated as an exosphere. Following Chamberlain, one may try to apply this formalism to the solar corona and interplanetary gas. From an experimental point of view an upper limit is obtained from the zodiacal light observations which are accounted for mainly on the basis of interplanetary dust.

The distribution of interplanetary dust is of particular interest in the vicinity of planets. It turns out that the concentration increases and reaches a maximum close to the surface of a planet, e.g. the Earth. One may also deduce the counting rates to be expected with a micrometeor detector. The comparison with observations leads to conclusions concerning the origin of the dust particles.

THE STATE OF MOTION OF THE INTERPLANETARY PLASMA

LUDWIG BIERMANN
Max-Planck Institut fur Physik und Astrophysik
Munchen, Germany

Abstract

The strong interaction between the plasma tails of comets (Type I in the usual classification) and the solar corpuscular radiation has the consequence that these tails can effectively be considered as probes for the state of motion and the particle intensity of the interplanetary plasma. This concept is supported by the correlation of special events in the comets' tails with geomagnetic phenomena. On this basis conclusions regarding the nature and the state of motion of the interplanetary plasma can be drawn, in particular the identity of the interplanetary plasma with the solar corpuscular radiation can be established. The several components of the solar corpuscular radiation known from geomagnetic observations can thus be studied by means of the comets' tails, without restriction to the vicinity of the ecliptic plane. Finally possible experiments will be discussed, in which plasma clouds emitted from space probes act as probes for the interplanetary space, in analogy to the plasma tails of comets.

References

1. *Der Schweif Des Kometen Mrkos 1957 To Appear:* Z. F. Astrophysik (1961 or 62) –Rhea Lüst.
2. Bartels, J., *Magnetische Charakterzahlen und Kennziffern* (1932/33 und 1940-1950).
3. Bartels, J., *Planetary Geomagnetic Activity in Graphic Representation* (Daily Cp, 1937-1958 und three hourly KP, 1937-1939, 1950-1958) ABHANDLUNGEN Der Akademie der Wissenschaften in Gottingen Mathematisch-Physikalische Klasse Heft 3.
4. Zeitschrift für Astrophysik 49, 111-119 (1960) Aus dem Max-Planck-Institut für Physik und Astrophysik, München. *Zur Enstehung der CO+ -Ionen in Kometen* Von L. Biermann and E. Trefftz.
5. Sonderdruck aus ASTRONOMISCHE NACHRICHTEN Band 286 Heft 2. *Zur Korrelation zwischen dem Auftreten von Kometenschweifen des Tpy 1 und der solaren Korpuskularstrahlung* Von P. Stumpff, Heidelberg.
6. *Kometenschweife und solare Korpuskular-Strahlung zu Zeiten geringer Sonnenaktivität* Ludwig Biermann, Reimar Lust and Eleanore Trefftz, Max-Planck-Institut fur Physik und Astrophysik, München.
7. *Zur Untersuchung des interplanetaren Mediums mit Hilfe kunstlich eingebrachter Ionenwolken* L. Biermann, R. Lust, Rhea Lüst und H. U. Schmidt. Max Planck-Institut fur Physik und Astrophysik, München.
8. Extrait des Memoires de la Societe Royale des Sciences de Liege Quatrieme Serie, Tome XIII, Fasc. 1-11(1953) 21–*Physical Processes in Comet Tails and Their Relation to Solar Activity* by L. Biermann Max-Planck Instut fur Physik, Gottingen Germany.

9. Reprinted from The Observatory, Vol. 77, No. 898 pp. 109-110 *Solar Corpuscular Radiation and the Interplanetary Gas*–L. Biermann.
10. Zeitschrift fur Astrophysik, Bd. 29, S.274-286 (1951) *Kometenschweife und solare Korpuskularstrahlung*, Von L. Biermann.
11. Zeitschrift fur Astrophysik, Bd. 47, S.225-265 (1959) *Die Bewegung der Materie im Schweif des Kometen Morehouse am 29. Oktober 1908*–Von P. Stumpff.
12. Zeitschrift fur Astrophysik 51, 163-176 (1961)–Aus dem Max Planck-Institut fur Physik und Astrophysik, Munchen *Aktivitat von Kometenschweifen in Perioden geomagnatischer Ruhe*–Von Rhea Lust.
13. Zeitschrift fur Astrophysik, Bd. 47, S.205-207 (1959) Kurze Mitteilung–*Die Bewegung von Schweifmaterie des Kometen Mrkos (1957d)*–Von Rhea Lust.
14. Sonderabdruck aus der ZEITSCHRIFT FUR NATURFORSCHUNG Band 7 a, Heft 1, 1952–Verlag der Zeitschrift fur Naturforschung, Tubingen, Mathildenstr. 29 *Uber den Schweif des Kometen Halley im Jahre 1910*–Von L. Biermann.

N63-10224

METEORITES: ORIGIN AND DISTRIBUTION

JOHN A. WOOD
Smithsonian Astrophysical Observatory

Meteorites are chunks of interplanetary matter which enter the earth's atmosphere at velocities of 15 to 20 km/sec. Friction with the atmosphere melts the surface of the object, and a bright light is emitted, which persists until the meteorite is decelerated to about 5 km/sec. By the time it has descended to an altitude of about 20 km, increasing air density sets up internal stresses in the meteorite which cause it to burst; the explosion is usually audible for many miles. The fragments then fall to the earth's surface, embedding themselves inches or feet deep in a farmer's field. Scientists come to the farm, and after several days of haggling, leave with the meteorite. The farmer then goes out and purchases a new car, not necessarily one of the "low-priced three".

Figure 1 illustrates relative proportions of the different types of meteorites. This graph is based only on those meteorites which are actually seen to fall. To include those found but not seen to fall, would bias our statistics in favor of those meteorites which are most radically different from ordinary terrestrial rocks, and especially the class of iron meteorites. Taken on the basis of falls only, however, it can be seen that most meteoritic material, over 90%, is stony in character. Some of these stony meteorites are unmistakably igneous in character. Most of them, however are chondrites, which are unlike any terrestrial type of rock; more about these later.

Most of us have the impression, perhaps gained in junior high school general science class, that meteorites are vast jagged hunks of nickel-iron. Flashy meteorites of this type do exist, but they make up a small minority of falls. Most meteorites are stones, and quite unspectacular stones at that, looking more like concrete than anything else. Of course, statistics based on falls are not impeccable; the meteorites must take quite a knocking around in space and in passage through the atmosphere, so the array of meteorite falls we study must be biased in favor of the physically stronger types of meteorite. But they're the best we have.

How often do meteorites encounter the earth? Hawkins (1) estimates the following influx law:

$$\log_{10} N = -\log_{10} m \quad -7.5 \text{ (per km}^2 \text{ hour)},$$

where N is the rate of influx of meteorites having mass greater than or equal to m (in kilograms, before entering the atmosphere). This influx law, which is based on the rate of arrival of new meteorites over the years, predicts that several hundred meteorites of mass greater than a kilogram must bombard the earth every year.

THE COMPOSITION, ATOMIC REACTIONS, AND AIRGLOW OF THE UPPER ATMOSPHERE

CHARLES A. BARTH
California Institute of Technology
Pasadena, California

Abstract

The physical structure of the earth's atmosphere is the result of the combined effect of the earth's gravitational field and the direct and indirect absorption of energy from the sun. The chemical composition in the mesosphere is determined by the photodecomposition of atmospheric gases by solar radiation and their recombination through atomic reactions. The composition of the thermosphere is also affected by mixing and diffusion processes. Many of the upper atmosphere atomic reactions produce the atomic and molecular spectra of the night airglow. Similar reactions emit a nightglow in the upper atmospheres of Mars and Venus.

References

1. Charles A. Barth, "Nitrogen and Oxygen Reactions in the Chemosphere" Proc. of the Conference on Chemical Reactions in the Lower and Upper Atmosphere", Interscience Publishers, 1961.
2. J. A. Radcliff, ed., "Physics of the Upper Atmosphere", Acadamic Press, 1960.
3. J. W. Chamberlain, "Physics of the Aurora and Airglow", Academic Press, 1961.
4. C. A. Barth and A. F. Hildebrandt, "The 5577A Airglow Emission Mechanism", Jour. of Geophysical Research *66*, 985 (1961).
5. C. A. Barth and J. Kaplan, "The Ultraviolet Spectrum of the Oxygen Afterglow and the Night Airglow", Jour. of Molecular Spectroscopy *3*, 583 (1959).

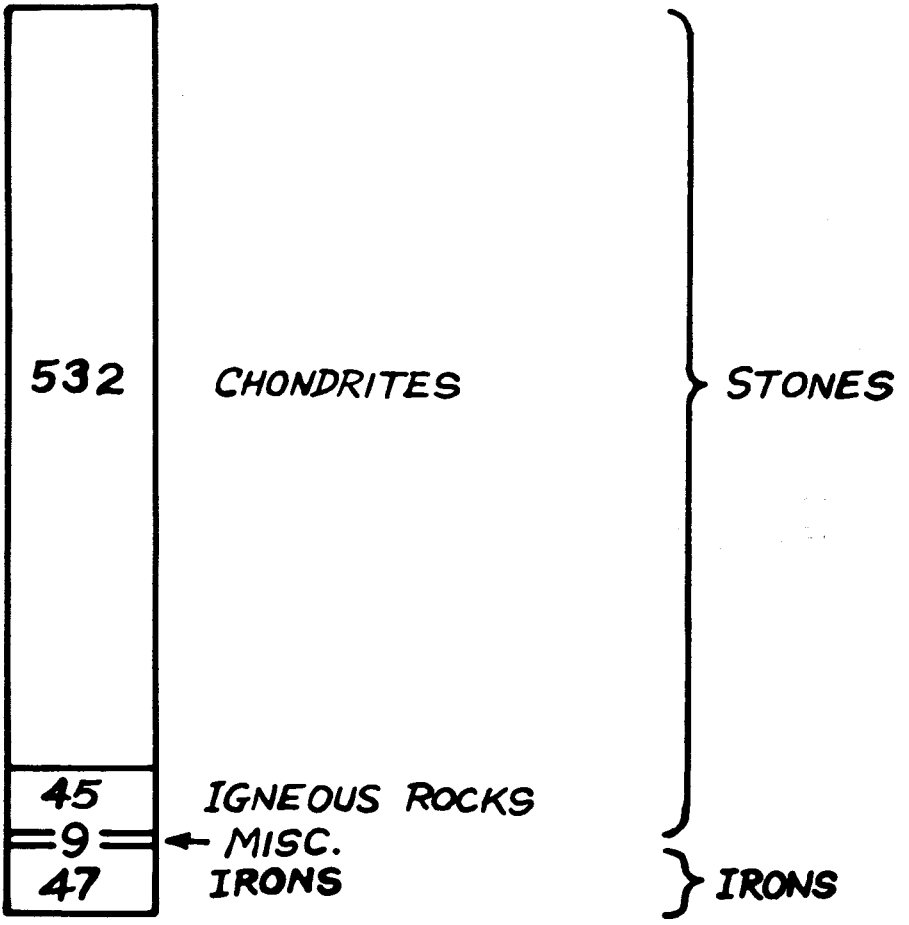


Fig. 1—Relative number of meteorites of different classes actually seen to fall to earth. Compiled from Prior and Hey's *Catalogue of Meteorites* (25).

It would be interesting to know something about fall frequency as a function of position on the earth, that is, as a function of latitude; but obviously this problem is hopeless, because of the non-uniform distribution of land area, rural population density, and other factors, over the earth. As an example of the variability of meteorite collection efficiency, records show that as of 1953, 102 meteorites had been collected in India and Pakistan, while only 8 were collected in China. Apparently India under British colonial administration was an ideal meteorite collecting net.

Unlike the meteor showers, which are associated with comet orbits, meteorites show no apparent preference to fall at one or more particular times of the year. More meteorites are collected during the summer, but there is little doubt that this is only a result of longer days and more work-

ers in the fields during the summer in the Northern hemisphere. For obvious reasons, few meteorites are collected in the Southern hemisphere.

The next logical question to pose is, where does this cosmic rock come from? The consensus at present is that it comes from the asteroid belt, and that asteroids and meteorites are essentially identical, the only difference being that of size. I'll present five reasons for thinking this.

First, there's not really anywhere else they could come from. A decade or two ago Ernst Öpik defended the notion that some of the meteorites had hyperbolic orbits, thus were visitors from interstellar space; and two years ago Harold Urey proposed that some were fragments of the moon's surface, blasted loose by impacting asteroids. But both these people have since abandoned their schemes. The lunar origin postulated by Urey is now a popular way of looking at the history of tektites.

Second, the orbits of thousands of optically visible asteroids are known, and of these, eight or more pursue orbits which carry them inside the earth's orbit. One of these, Hermes, is capable of coming within 800,000 km of the earth, barely twice the distance away of the moon. It is logical to suppose these bright asteroids are accompanied by a great number of lesser bodies, too small to appear on our astronomical telescope plates. Some of these many asteroids must inevitably collide with the earth. Ergo, since by definition everything which falls from the heavens is a meteorite, at least some meteorites must be of asteroidal origin.

Third. As we will see later, many meteorites contain measurable amounts of noble gases: He, A, Ne, Xe. Goles, Fish and Anders (2) have examined in some detail the question of what average temperature meteorites would have in space, for different types of orbits, and how effective this temperature would be in "cooking out" a meteorite's supply of noble gases. They found that in order for meteorites to retain noble gases, as they have, their mean environmental temperature can have been no higher than about 200° K. 200° K is the temperature which would obtain in a black body whose orbital mean distance is about 1.4 A.U.* This rules out Urey's lunar hypothesis, mentioned above, because meteorites derived from the moon could be expected to pursue orbits with mean distance near one A.U.; that is, earth-like orbits. The temperature this close to the sun would be too high to permit the meteorites to retain noble gases.

Fourth. An argument I've worked on is based on the statistics of meteorite falls vs. time of day (3). Figure 2 is a histogram showing the total numbers of meteorites seen to fall during each hour of daylight. Obviously, substantially more fall in the afternoon than in the morning. My argument turns upon whether these statistics are valid or not. Harold Urey has asserted that the poor showing in the mornings is because less people are up and about and working the fields in the mornings than in the afternoons. I admit this factor must bias the histogram, but I don't think it alone could produce such a striking concentration of falls in the afternoon. I might point out here that Virginia Polytechnic Institute would probably be uniquely qualified to pronounce on this point of statistics of rural work patterns. If you

*An Astronomical Unit (A.U.) is the mean radius of the earth's orbit.

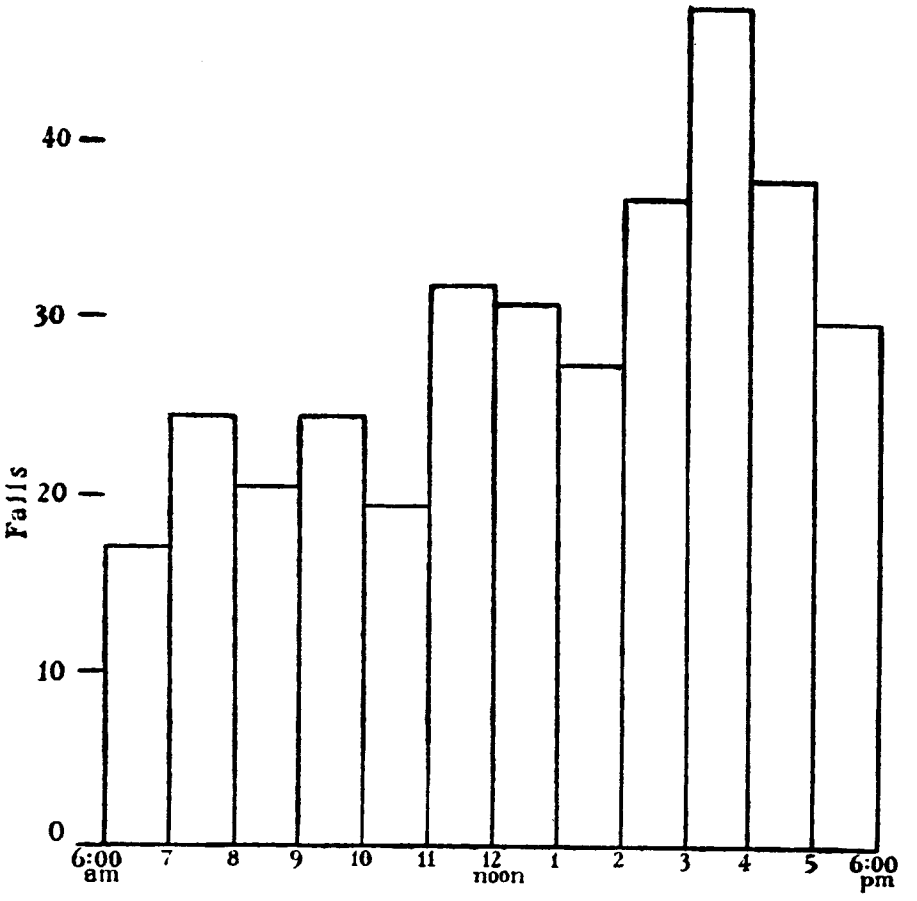


Fig. 2—Histogram showing number of meteorites which have been seen to fall and reported during each hour of daylight. From *Monthly Notices of the Royal Astronomical Society*, Vol. 122, p. 79, 1961.

grant that the statistics are meaningful, then the sketch in Figure 3 shows in a nutshell what they mean. Here we're looking down on the plane of the ecliptic. The earth's rotation being counterclockwise, you can see that the AM hours on the earth always constitute the earth's leading edge, so to speak, and the PM hours its trailing edge. Then, a preponderance of meteorite falls in the afternoon means that most meteorites tend to overtake the earth from behind. A faster orbital velocity than the earth's means an eliptical orbit, with aphelion outside the earth's orbit. According to my calculations, most of these orbits would be similar to asteroid orbits.

Fifth, it would clinch the matter if we actually knew the orbits of some of the meteorites which have fallen. A number of orbits have been calculated from the reports of carefully interviewed witnesses of falls, but the amount

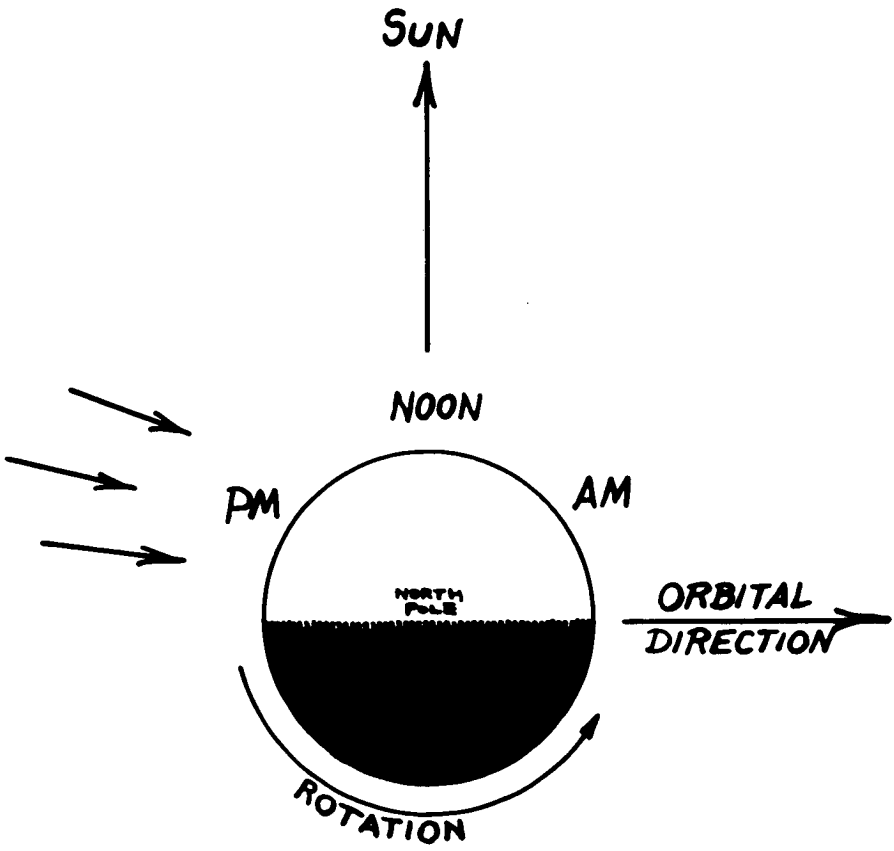


Fig. 3—Sketch showing relationship of morning and afternoon to orientation of earth in space; direction of approach of "afternoon meteorites".

of error inherent in reports by lay witnesses makes these orbits at best doubtful. However, we now have one good meteorite orbit. This chondrite fell at Přibriam, Czechoslovakia (4), and by accident was recorded on astronomical plates at two different observatories. Given two-station photography, it was possible to calculate an orbit with considerable precision. Figure 4 shows the shape of the orbit, and lists the orbital elements. As you see, the meteorite reached aphelion quite far out, close to Jupiter's orbit. The asteroid belt lies between Mars and Jupiter. Přibiam's was definitely an asteroid-like orbit. Its orbital plane was inclined to the plane of the ecliptic by some 10 degrees, the intersection of the two planes being denoted in Figure 4 by a dashed line. The orbits of earth and meteorite coincide at the meteorite orbit's descending node.

Hoping to be able to derive more meteorite orbits, the Smithsonian Astrophysical Observatory plans to set up a network of thirteen automatic camera stations in the midwest (Figure 5), which will continuously photograph the

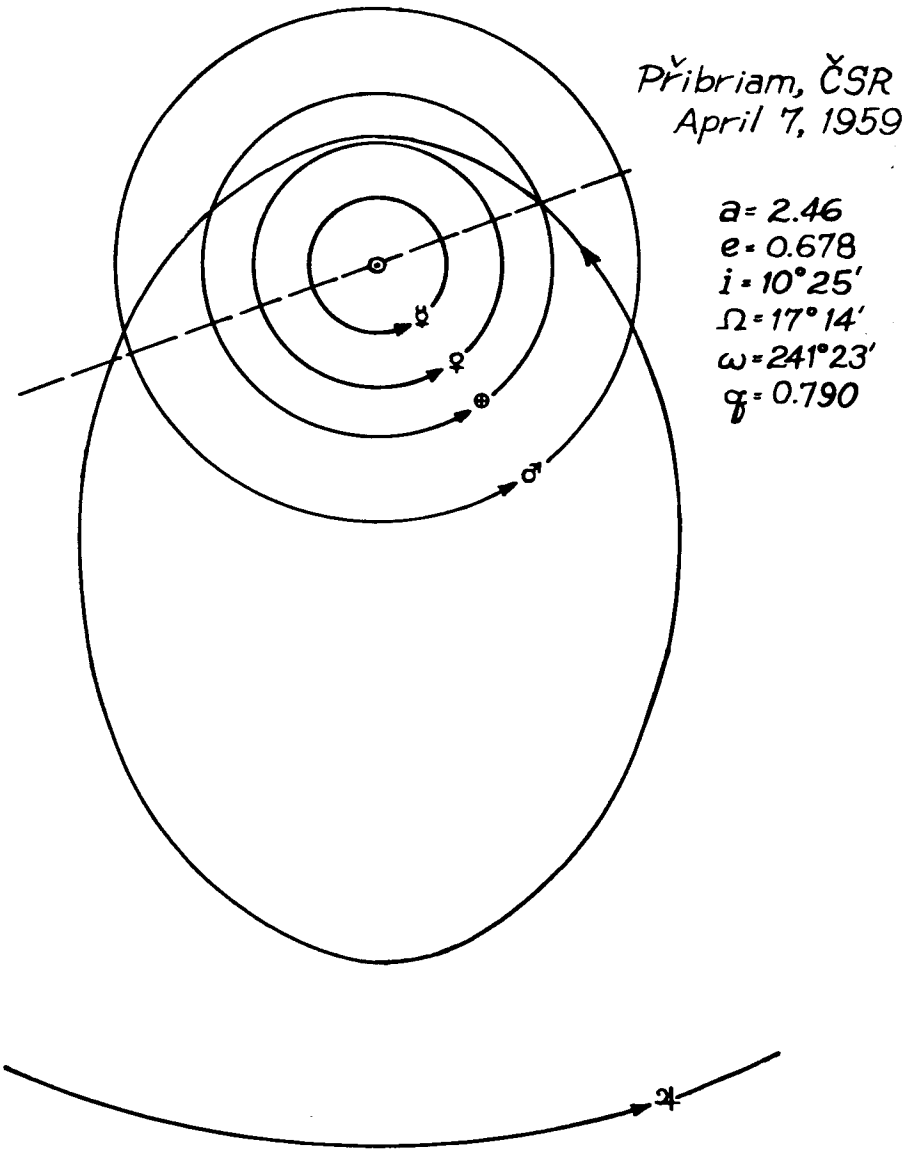


Fig. 4—Obit and orbital elements of the Pribriam meteorite, the only meteorite fall to be photographed in flight from two stations (4). Dashed line: intersection of the orbital plane with plane of the ecliptic; angular separation of the two planes is 10° 25′. The arrowhead is closest to the descending node and point of collision.

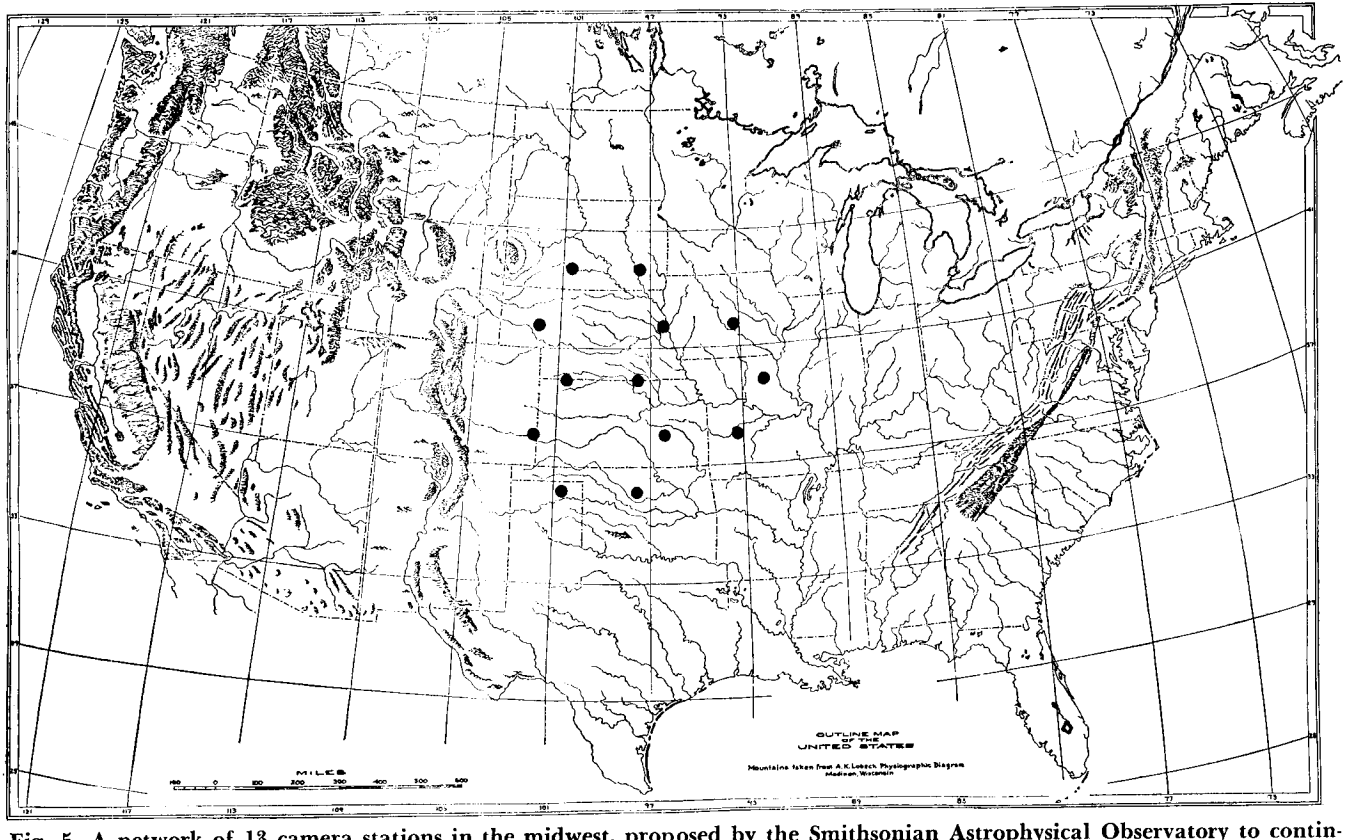


Fig. 5—A network of 13 camera stations in the midwest, proposed by the Smithsonian Astrophysical Observatory to continuously photograph the night sky, yielding multiple-station photographs of an anticipated 1 to 5 meteorites per year.

night sky. Any meteorite which falls within this honeycomb pattern at night will appear on photographs from at least two stations. Richard McCrosky, who is engineering this scheme, anticipates that it will catch from one to five meteorites larger than a kilogram, per year. The photographs will yield not only orbits, but will permit us to locate the fallen meteorite itself to within 500 meters, so that it can be recovered as promptly as possible.

In the absence of any statistical information on meteorite orbits, and having attempted to demonstrate that meteorites are probably identical with asteroids, perhaps it would be worth looking briefly at the distribution of asteroids. Figure 6 shows the asteroids known in 1887 (5), which was already quite a few, sorted out into a histogram according to the semi-major axis (or mean distance) of the orbit each follows. Plotted on this line you see the position of the Sun, Mercury, Venus, Earth, Mars, and Jupiter. In 1887, all the asteroids fell between Mars and Jupiter, in the so-called "asteroid belt", but today we know of several whose mean distances fall inside of Mars' orbit. The rational fractions which appear under the histogram refer to the period which an asteroid of that particular mean distance would have, relative to the period of Jupiter. These particular orbits "resonate" with Jupiter's orbit. Repeated perturbation by Jupiter would make these orbits unstable, and for this reason we find corresponding gaps, known as Kirkwood's gaps, in the asteroid distribution.

In Figure 7, we see the result of an extensive survey of asteroids and their brightness by Gerard Kuiper, then at the Yerkes Observatory, and his group (6). The top curve represents all the asteroids they observed. Here the

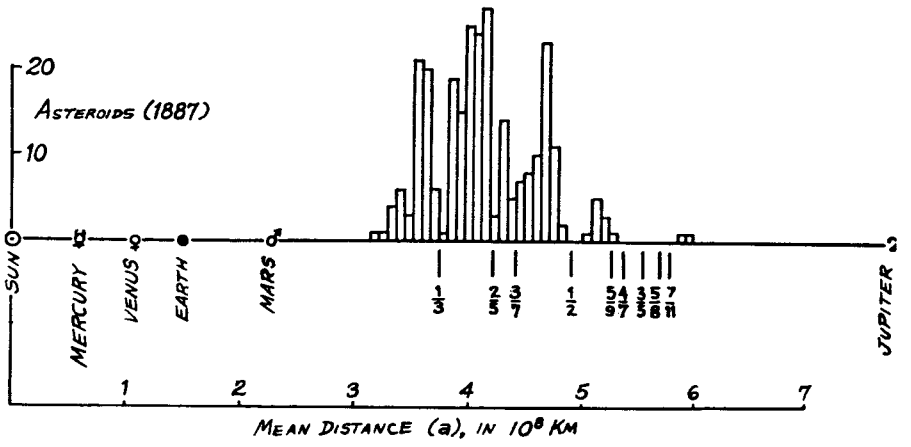


Fig. 6—A histogram of asteroids known in 1887 and their mean distances (i.e., semi-major axes of their orbits). The rational fractions refer to the period which an asteroid of that particular mean distance would have, relative to the period of Jupiter. These orbits "resonate" with Jupiter's orbit, and perturbations by the latter cause the gaps ("Kirkwood's Gaps") in the asteroid distribution at 1/3, 2/5, 3/7, etc. of Jupiter's period. Compiled from *The Asteroids* by D. Kirkwood (5).

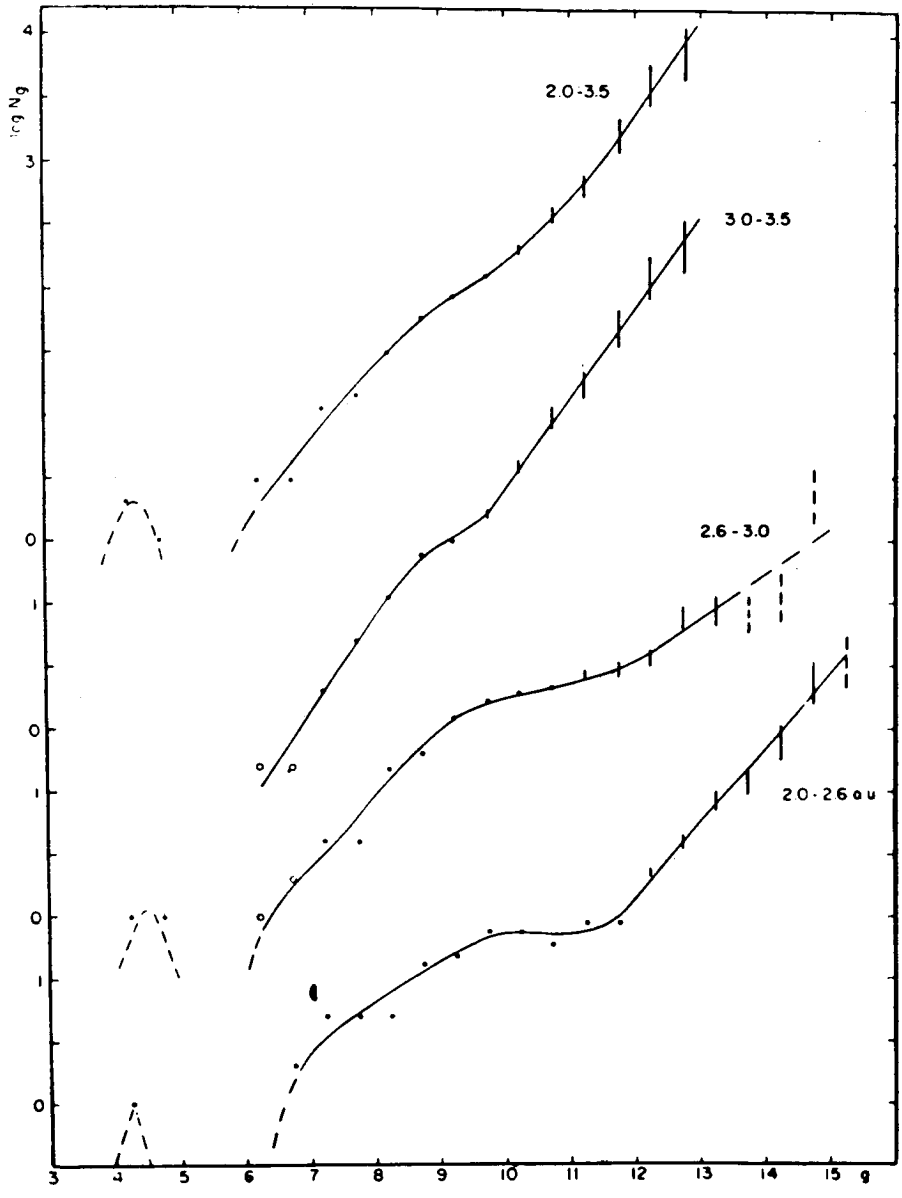


Fig. 7—Frequency distribution of observed asteroids (ordinate is log frequency) vs. absolute magnitudes (abscissa) for three distance zones and their sum. From Kuiper, et al, "Survey of Asteroids", *Astrophysical Journal Supplement* Ser. 3, Fig. 5. Copyright 1958 by the University of Chicago.

ordinate is log of frequency, and the abscissa is absolute magnitude, which is proportional to the square of the radius of an asteroid. Kuiper notes that the occurrence of the three bright asteroids (Ceres, Pallas and Vesta; these appear on a small separate peak in the left of Fig. 7) seems to be a separate phenomenon, somewhat outside the smooth frequency curve representing the normal asteroids. He would like to think these three represent original, untampered-with planetary condensations, while the rest are fragments of other condensations which have been disrupted.

If you're satisfied that meteorites seem to come from the asteroid belt, the next logical question is where the asteroids came from. The most popular answer is what I've just suggested, that they and the meteorites are fragments of a destroyed planet or set of planets which used to orbit between Mars and Jupiter. There are several reasons for believing this.

For one, the brightness of many of the asteroids is variable (7). Figure 8 shows one of the more spectacular examples, 44 Nysa. Relative visual magnitude is plotted against time. Twenty-six such variable asteroids have been investigated. This variable brightness is interpreted to mean that these asteroids are non-spherical, being rather oblong or flat in shape, so that in rotating they present a variable surface area to reflect the sun's rays. According to this concept, 44 Nysa takes about 6 hours to make a complete revolution. Making one more extension, we reason that if an asteroid is irregular in shape, then it is probably fragmental.

The other reason for believing in "parent meteorite planets", as Urey has termed them, depends on the internal properties of the meteorites themselves. Suffice it to say at this point that some of them contain evidence of having been subjected to very high temperatures, followed by millions of years of gradual, quiescent cooling. There also appears to have been a gravitational differentiation of a heavy liquid iron phase from a lighter silicate melt. It does not seem possible to understand these processes taking place other than in the body of a planet.

How did this parent meteorite planet become disrupted? It has not been possible to justify any kind of internal explosion, so the popular concept now is that there were several such planets, and that they collided with one another. The more planets you postulate, the more probable such an event becomes. Given 5 to 10 such planets, and reasonal initial conditions, Kuiper (8) estimates a probability of about 0.1 that a collision would occur in three billion years.

Another possibility is that one of these parent meteorite planets passed within the Roche limit of Jupiter. The Roche limit is a theoretical surface surrounding a massive body; a smaller body passing within the Roche limit would have such severe tidal stresses applied to it that it would be disrupted. This rather dramatic sketch (Figure 9) demonstrates the size of the Roche limit of Jupiter. This concept can be misused, however. Roche's derivation holds only for an ideal fluid body, so unless our parent meteorite planet was large enough to behave essentially like a fluid, it looks like we have to depend on mutual collisions to break them up.

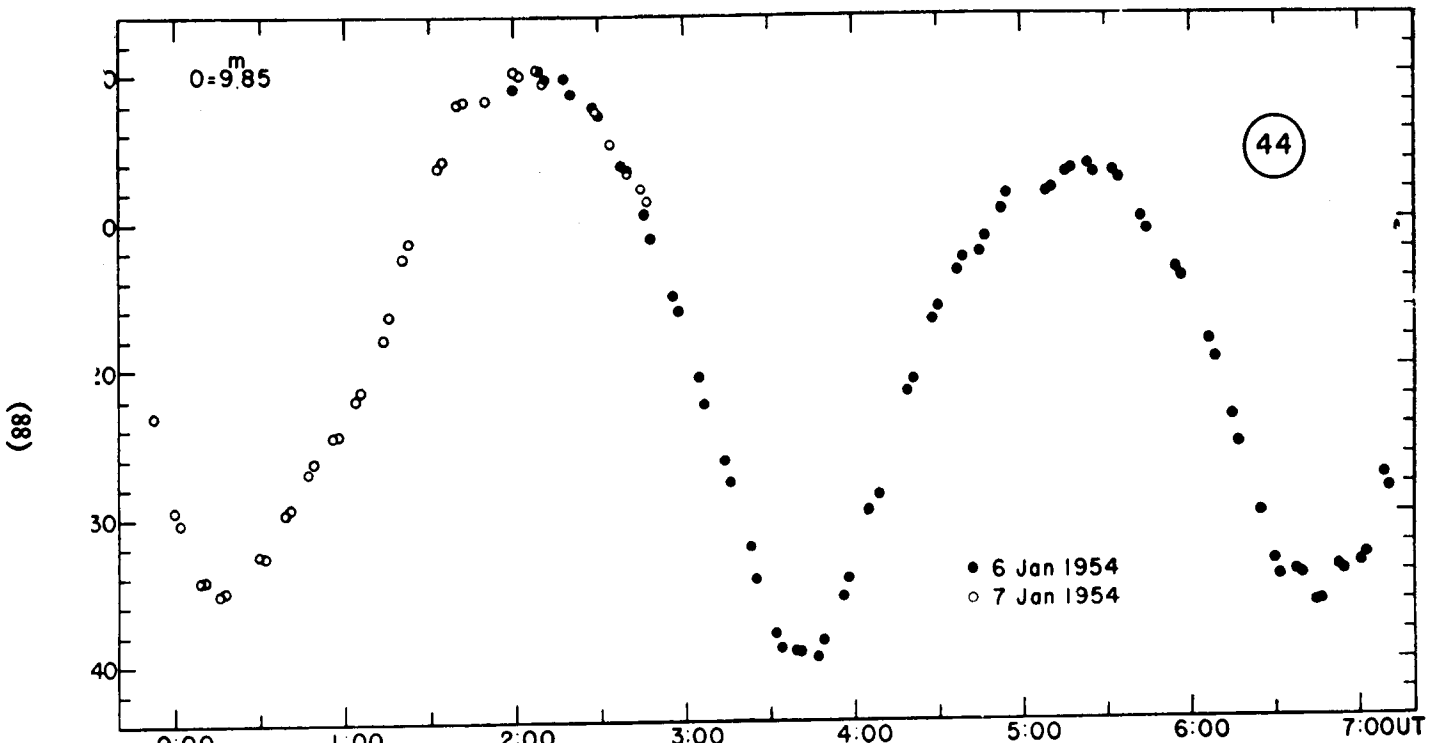


Fig. 8—Light curve of 44 Nysa, an asteroid of variable brightness. Ordinate, brightness (relative visual magnitude); abscissa, time in hours. From Groeneveld and Kuiper, "Photometric Studies of the Asteroids. II.", *Astrophysical Journal* 120, Fig. 12. Copyright 1954 by the University of Chicago.

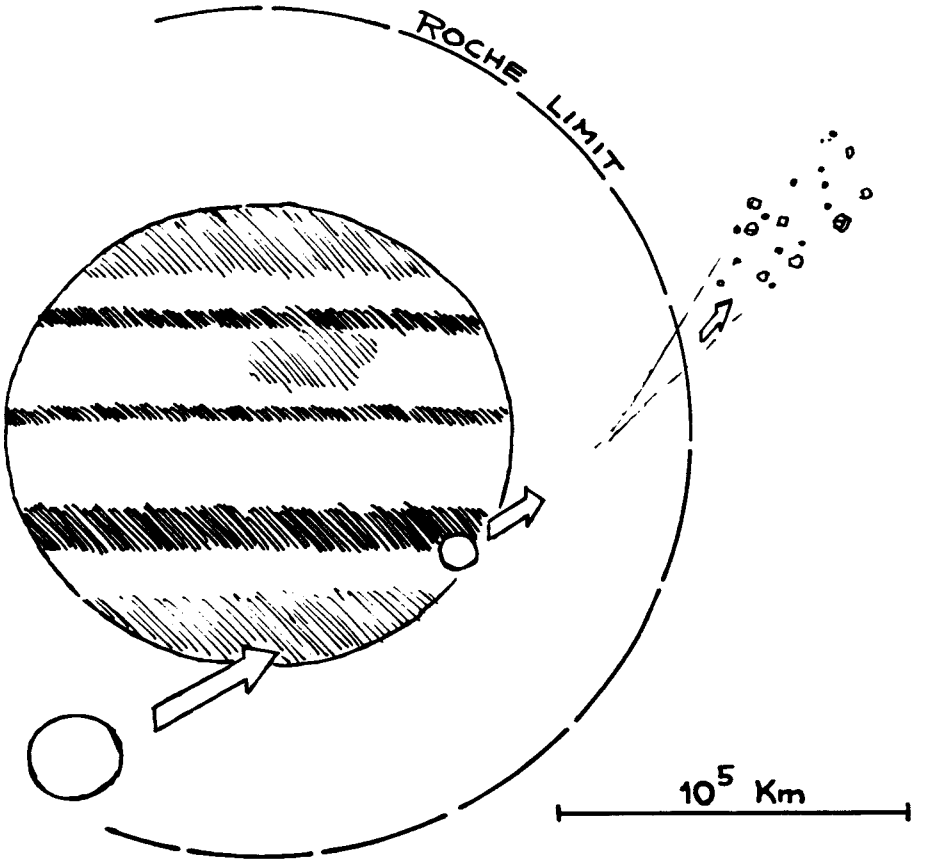


Fig. 9—Sketch showing the approximate dimension of Jupiter's Roche limit. A (fluid) object passing within this zone would be disrupted by tidal stresses.

So far I've talked about a consensus, and about "most people think", "it is generally agreed". If we inquire any further into the question of the history of the meteorites and parent meteorite planet, we leave the land of genial agreement, and embark on the stormy sea of violent controversy.

For example, when did this collision take place? There's no clear-cut answer yet, in spite of all the isotopic dating of meteorites that's been done. The stones have been retaining noble gases generated by cosmic ray impacts for tens of millions of years, the irons for hundreds of millions of years. Not until after collision could these meteorites have been exposed to cosmic rays. On the other hand, the stones have been accumulating radiogenic argon for billions of years. In the interior of a planet, temperatures developed by disintegrating U, Th and K^{40} would presumably be too high to permit the retention of argon. So perhaps these long radiogenic argon ages date the disruption of the parent meteorite planet? At this point, you can write your own history.

Another question you won't get a satisfactory answer to: How big were these parent meteorite planets? Harold Urey (9) and also A. E. Ringwood (10), in Australia, believe that it or they were quite large, probably bigger than the moon. Their compelling argument is that some meteorites, especially irons, contain diamonds. To form diamonds under equilibrium conditions requires high pressure, which in turn means a large planet.

But Gordon MacDonald (11) has pointed out that pressures adequate to form diamond are also more than adequate to form other conspicuous high-pressure minerals in the stony meteorites, namely the minerals found in eclogite rock on the earth. Terrestrial eclogites occur mostly in diamond pipes, where they were apparently carried up from great depth along with the diamonds. Many geophysicists believe that the crust of the earth is underlain by a zone of eclogite. The pressure-temperature stability fields of some of the eclogite minerals is fairly well known.

The fact is, though, that eclogite minerals have not formed in any of the stony meteorites, in spite of the fact that most of them have chemical compositions which would permit the appearance of these phases. Most students of meteoritics are coming around to the view that meteorite diamonds are not the product of equilibrium conditions at all, but were formed by the momentary high pressures attending high-velocity collision, either of parent meteorite planets with one another, or with the earth. This is an idea put forward by H. H. Nininger (12) years ago, and it has been vindicated this year by DeCarli and Jamieson (13), who succeeded in producing diamond in a pure graphite sample by subjecting it to an explosive shock, producing a maximum pressure estimated at 300,000 atmospheres.

Opposed to Urey and Ringwood, Fish, Goles and Anders (14) believe the parent planets were quite small, between 10 and 200 km in radius. U, Th and K^{40} would not be able to heat such small planets substantially, and so radiogenic argon could accumulate in them. In this case, the $K^{40}-A^{40}$ dates of meteorites have nothing to do with the time of collision and breakup.

It might be worth asking, before we go further into their internal properties, why we're so interested in meteorites. As I see it, there are two reasons. For one, if these meteorites are fragments of a destroyed planet, as most of us believe, then here we have a chance, probably the only chance, to sample and study the interior of a planet. After all, geologists are only able to study the upper one-tenth of a percent of the earth's radius at first hand. Our understanding of what goes on below that will probably always depend on theoretical considerations and analogies drawn with meteorites. One such analogy, venerable with age, assumes that the earth has a core of metallic nickel-iron, because of the occurrence of nickel-iron meteorites.

Obviously the analogy isn't very close, and comparisons must be drawn with care, because of the vast difference in sizes, and the relatively small role pressure seems to have played in the parent meteorite bodies.

The other reason is that preserved among the meteorites may well be unaltered samples of the original rock which first condensed or accreted during the birth of the solar system. To be able to study such material, I consider the most exciting possibility in earth sciences. There are good

reasons to expect to find such material. For one, the meteorites are old—very old. Much older than any terrestrial rock, on the basis of various isotopic dating methods. For another, much of the reworking and destruction of the original earth rock has been the result of weathering and sedimentary processes. These depend on the presence of water, organic life, and a hostile atmosphere. None of these agents were present on the parent meteorite planet. Further, these parent planets, being microscopic in size relative to the earth, have had the capacity to develop only a fraction of the internal temperature the earth was capable of. It is therefore that much more likely that portions of the parent meteorite planets survived destruction by igneous processes.

With that justification, I'll first review briefly results obtained in the most popular and populous branch of meteoritics, namely isotopic dating. There are three major techniques which are said to yield the "age" of meteorites. In each case, it is worth looking in some detail at just what the date means. The first of these is the Pb^{206}—Pb^{207} method. Pb^{206} is a product of U^{238} decay, and Pb^{207} of U^{235} decay. It is necessary to make measurements of the lead isotope ratios Pb^{207}/Pb^{204} and Pb^{206}/Pb^{204} in two or more different types of meteorite which have substantially different U/Pb ratios, as for example a chondrite and an iron meteorite. A knowledge of absolute amounts of lead or uranium or their ratio is not needed in making the calculations, however. What is done, in effect, is to project the lead isotope ratios of the several kinds of meteorites back in time until they are identical. The date we obtain is the time since all types of meteorites had the same kind of lead in them. Clair Patterson, who makes the assumption that the different meteorites types formed simultaneously and that each has since constituted a separate, closed system, concludes that his Pb^{206}—Pb^{207} ages date the actual origin of the meteorites (15). But if we postulate that the parent meteorite planets were initially homogeneous in composition, and subsequently were heated and partially melted, and that phase differentiation in the molten material produced the various different kinds of meteorites, then the Pb^{206}—Pb^{207} method dates that differentiation process, not the actual origin of the planet.

The Rb^{87}—Sr^{87} technique is in some respects similar to the Pb^{206}—Pb^{207} method. Rb^{87} decays into Sr^{87}. Measurements are made of the Sr^{87}/Sr^{86} ratio and of the absolute amounts of Sr and Rb or of their ratio, again in several types of meteorite. Since iron meteorites are virtually devoid of these elements, the pair selected has always been one chondrite and one of the igneous types of stony meteorite. As with the Pb^{206}—Pb^{207} method, the computation amounts to projecting back in time until both stones had strontium of the same composition. The interpretation of the date yielded is the same as the interpretation of the Pb^{206}—Pb^{207} number; time since simultaneous origin of the two meteorite types, or time since melting and phase fractionation took place.

Finally, of the three, the largest number of measurements have been made by the K^{40}—A^{40} method. Here one measures the absolute amount of potassium (K^{40} is the parent isotope) and the amount and isotope composi-

tion of argon (A^{40} is a daughter isotope). A straightforward calculation then yields the time that was necessary for K^{40} decay to produce the observed amount of A^{40}. As I mentioned earlier, heat or even warmth causes these stones to "leak" argon, so the meaning of the K^{40} —A^{40} date is simply the period of time since they became quite cold; Goles, Fish and Anders (2) estimate a maximum temperature of about 200° K.

U^{235}, U^{238}, Rb^{67} and K^{40} have long half-lives, appropriate for trying to measure the age of the solar system. Intriguingly enough, all three techniques yield dates which cluster about the same value: 4.5×10^9 years (15, 16, 17).

An entirely different type of isotopic date is the so-called "cosmogenic" age. Here an attempt is made to measure the length of time a given meteorite has been exposed to cosmic ray bombardment in space. Cosmic rays striking heavy elements in the meteorites generate certain spallation products; the spallation isotopes usually extracted and measured are the noble gases, especially He^3, He^4, tritium, Ne^{21} and A^{38}. Making certain assumptions about the energy spectrum of cosmic rays, the production rates of these same isotopes in appropriate target materials placed in particle accelerators can be measured. This production rate, plus the isotope abundance in a given meteorite are used to calculate that meteorite's exposure age.

Speaking in generalizations, the stony meteorites tend to have exposure ages in the tens of millions of years, and the iron meteorites in the hundreds of millions of years. Many workers now feel that these numbers do not date the original collision between parent meteorite bodies. After all, the fragments from the original collision probably undergo second-order, third-order and *n*th-order collisions with one another in subsequent years. With each such collision, particle sizes are reduced, and new meteorite surface exposed. Thus, the cosmic ray exposure age of a meteorite merely dates its *last* collision, and since it is reasonable to picture a more or less steady bumping together and grinding up of bodies in the asteroid belt, we should expect to encounter a wide spectrum of exposure ages, which is just what has been found. The order of magnitude discrepancy between "average" stone and iron values probably reflects the superior toughness of iron meteorites, and the lesser frequency with which they are broken by collisions.

Figure 10 illustrates a rather interesting piece of work done by Fireman at the Smithsonian Astrophysical Observatory (18). He has sampled a slab of the iron meteorite Grant, and measured the content of the spallation isotope He^3 in each sample. The results are contoured in this figure. As is to be expected, concentration of He^3 decreases steadily with depth in the meteorite, as shielding by the overlying iron attenuated the cosmic ray flux. The shape of the contours is taken to reflect the shape the iron meteorite had in space, before it entered the earth's atmosphere and was probably broken during flight or impact.

Measurements of spallation isotopes in meteorites have also added to our knowledge of cosmic radiation. For example, comparison of the amount of a long half-life radioactive spallation product with a shorter half-life spallation product tells us something about the constancy of the cosmic ray flux

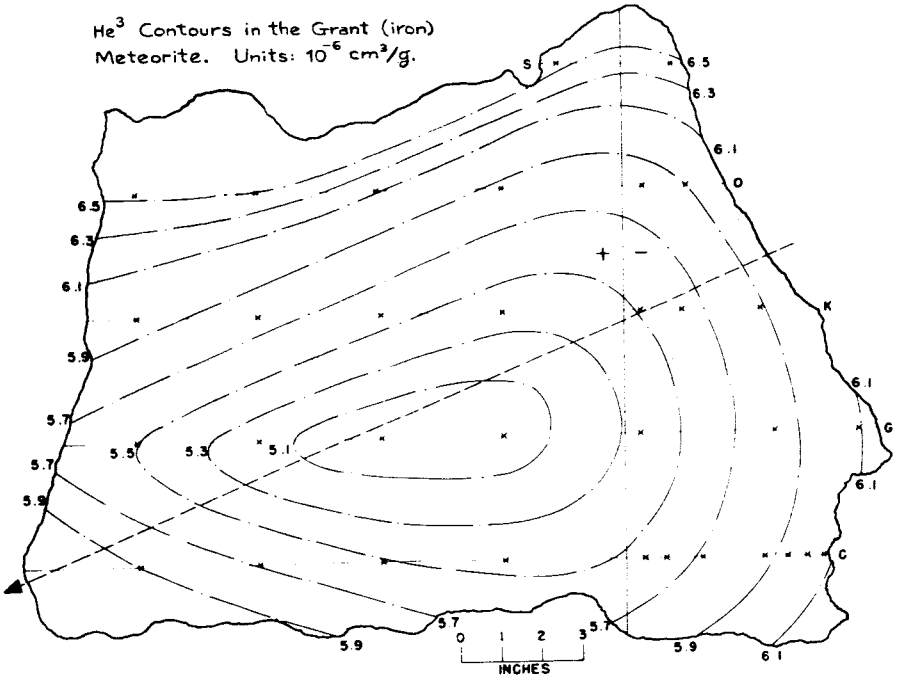


Fig. 10—Helium—3 contours in the Grant (iron) meteorite. Units are 10^{-6} cm^3/gm. Contours reflect the shape of the iron before entering the earth's asmosphere. From Fireman, *Planetary and Space Science* 1, p. 66, 1959.

with time. Again, if you can measure the amount of a radioactive spallation product whose half-life is short relative to the period of an asteroid, then this abundance will reflect the cosmic ray flux close to the earth; this can be compared with the abundance of a spallation isotope with longer half-life, which reflects the integrated flux intensity over the whole of the meteorite's orbit. The comparison should tell us something about the constancy of cosmic radiation in space. The short half-life isotope which has been measured is A^{37} (35 days). To make a measurement of this type requires fast action when a new meteorite falls, in order to get it into the laboratory before the A^{37} decays. This experiment was performed last year on the Hamlet meteorite, but results of different workers disagreed. Stoenner, Schaeffer and Davis (19) concluded that cosmic ray flux was probably uniform throughout the solar system, but Fireman and DeFelice (20) reported a somewhat higher cosmic ray intensity closer to the sun.

The latest grail of the meteorite isotope men has been the quest for so-called fossil radioactivity. But I feel that the meaning of this problem is closely connected with the question of the thermal history of the meteorites, so I propose to go into these together.

First, it is evident that heat has played an important part in the history of the parent meteorite planets. Mention was made earlier of igneous meteorites.

Figure 11 is a thin-section of the meteorite Stannern. It is viewed by transmitted, unpolarized light, as will be all subsequent photographs of meteorite thin-sections. Stannern is typical of the class of meteorites called eucrites. Eucrites are not uncommon, in fact they make up the bulk of the igneous types of meteorites seen to fall. These eucrites are virtually identical in mineralogy and texture to terrestrial diabases, igneous rocks most of which have flowed into cracks in cooler rock and hardened relatively quickly. The light (more transparent) mineral in the slide is labradorite, the dark mineral an augitic pyroxene. Eucrites are perhaps the most striking evidence of high temperatures in the parent meteorite planet.

Second, there has also been cooling. The iron meteorites probably differentiated as a liquid, at 1300°C or more, from molten silicates. Most of the iron meteorites now consist of collossal single crystals, a foot or several feet in diameter. Such crystals can have been produced only by very slow, steady cooling. The phase configuration in the iron meteorites allows an estimate to be made of the cooling rate, which I shall discuss shortly. Also relative to cooling, as we have already seen, K^{40} —A^{40} dates indicate that cooling was completed some 4.5 billion years ago.

One of the most important questions in meteoritics today is, where did this heat come from? It turns out to be difficult to answer, for there are a number of boundary conditions that the thermal history has to satisfy.

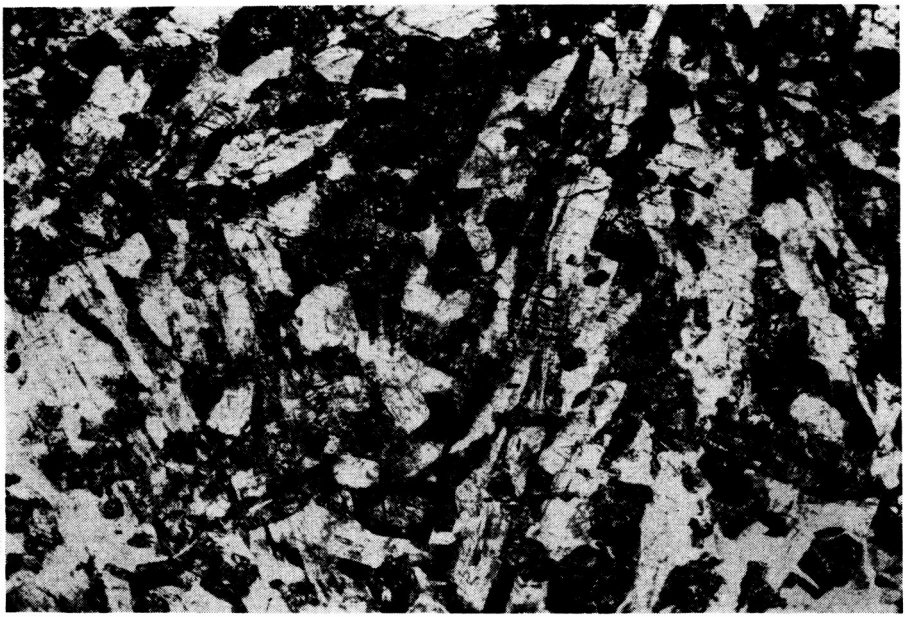


Fig. 11—Thin-section of the meteorite Stannern, viewed by transmitted unpolarized light (as are all subsequent thin-section photographs). Stannern is typical of the common eucritic type of meteorite. Magnification, 45 diameters.

A first attempt at answering it might be to say that the heat was generated by the disintegration of radioactive elements, as we know it is being generated in the earth's crust. The content of uranium, thorium and potassium has been measured in chondritic meteorites, and turns out to vary amazingly

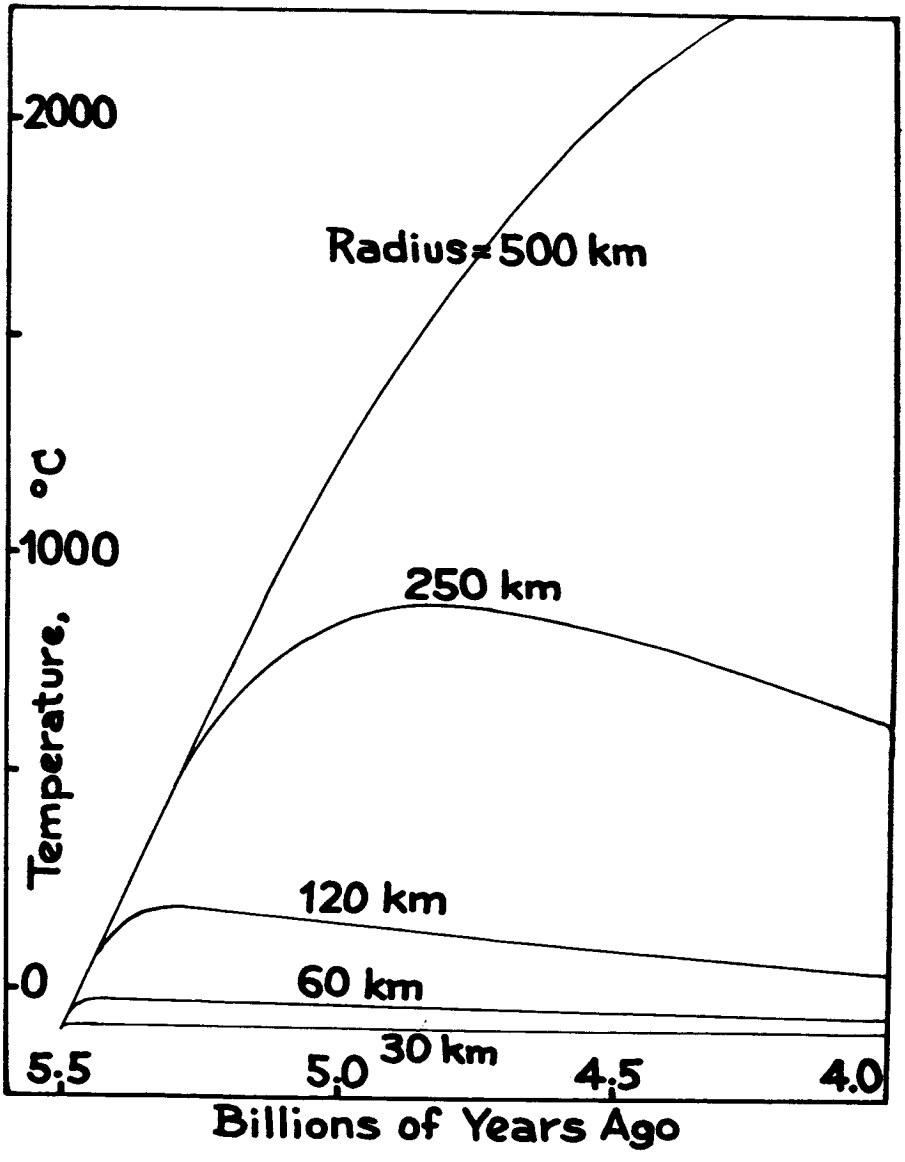


Fig. 12—Central temperature in asteroids of chondritic composition as a function of time and radius of the asteroid. The asteroids are assumed to have formed 5.5×10^9 years ago. Heat is generated by U, Th and K^{40} decay.

little from chondrite to chondrite. Using these values, it's easy to calculate the heating curve in asteroids of various sizes, assuming them to be chondritic in composition. The curves in Figure 12 represent temperature at the *center* of the asteroid. Assume for the sake of argument that the asteroids accreted and began heating up 5.5 billion years ago. It can be seen that the parent asteroid would have to have been larger than 250 km in radius in order to attain melting temperatures at its core. A radius of about 350 km would probably do the job nicely. But it is obvious that having once gotten up to melting temperature, the asteroid would be very slow about cooling down again—much too slow, since the chondrites must have been cold enough to retain A^{40} since 4.5 billion years ago.

There are two ways of getting around this difficulty. One is to propose that the initial collision and breakup took place just 4.5 billion years ago, and that the fragments then cooled quickly by radiation. The other is to conclude that U, Th and K^{40} are not the heat source, that there was another heat source, and that the planet was so small that U, Th and K^{40} did not contribute significantly to the heat balance. Edward Anders (14) has taken the latter course.

The elements which go to make up our solar system are thought to have been created by thermonuclear reactions in stellar interiors. While all the stable and radioactive elements which we find in our chart of the nuclides were being formed, who is to say that certain other radioactive elements having relatively short half-lives might not also have appeared? If planetary condensation followed rather quickly on the heels of nucleogenesis, some of these short half-life elements would be incorporated in the planets. If, in turn, these elements were abundant enough, they could heat the interior of an asteroid; maybe even a rather small asteroid. The most hopeful potential muclide would be Al^{26}, with half-life of about a million years. Then heating and cooling would take place very rapidly compared with U-Th-K^{40} heating, and there is no difficulty explaining cold asteroidal material by 4.5 billion years ago. Naturally there would be no Al^{26} left in meteorites today, hence the term "extinct radioactivity" applied to the concept. Nor could you expect to find the daughter of the Al^{26} decay, for this is Mg^{26}, and would be swamped out by primordial Mg^{26}, which is a common stable isotope.

This hypothesis of Anders' gained sudden impetus when daughters of short half-life isotopes actually began to be found in meteorites—not the daughter of Al^{26} or any potential heat-generator, but daughters of other extinct isotopes. The first such discovery was made by Reynolds (21).

The spectrometer trace Reynolds obtained for Xenon extracted from the chondrite Richardton is reproduced in Figure 13. The small horizontal lines show where the peak heights of terrestrial Xe would stand, if the two spectra are adjusted to have equal Xe^{132}. Though there are other discrepancies, the most conspicuous feature of the spectrum is the unexpectedly high Xe^{129} peak. Richardton's Xe is enriched in Xe^{129} by perhaps 50%. After examining other possible reasons for this anomaly, Reynolds concluded that it could only be due to production of radiogenic Xe^{129}. The parent of Xe^{129}

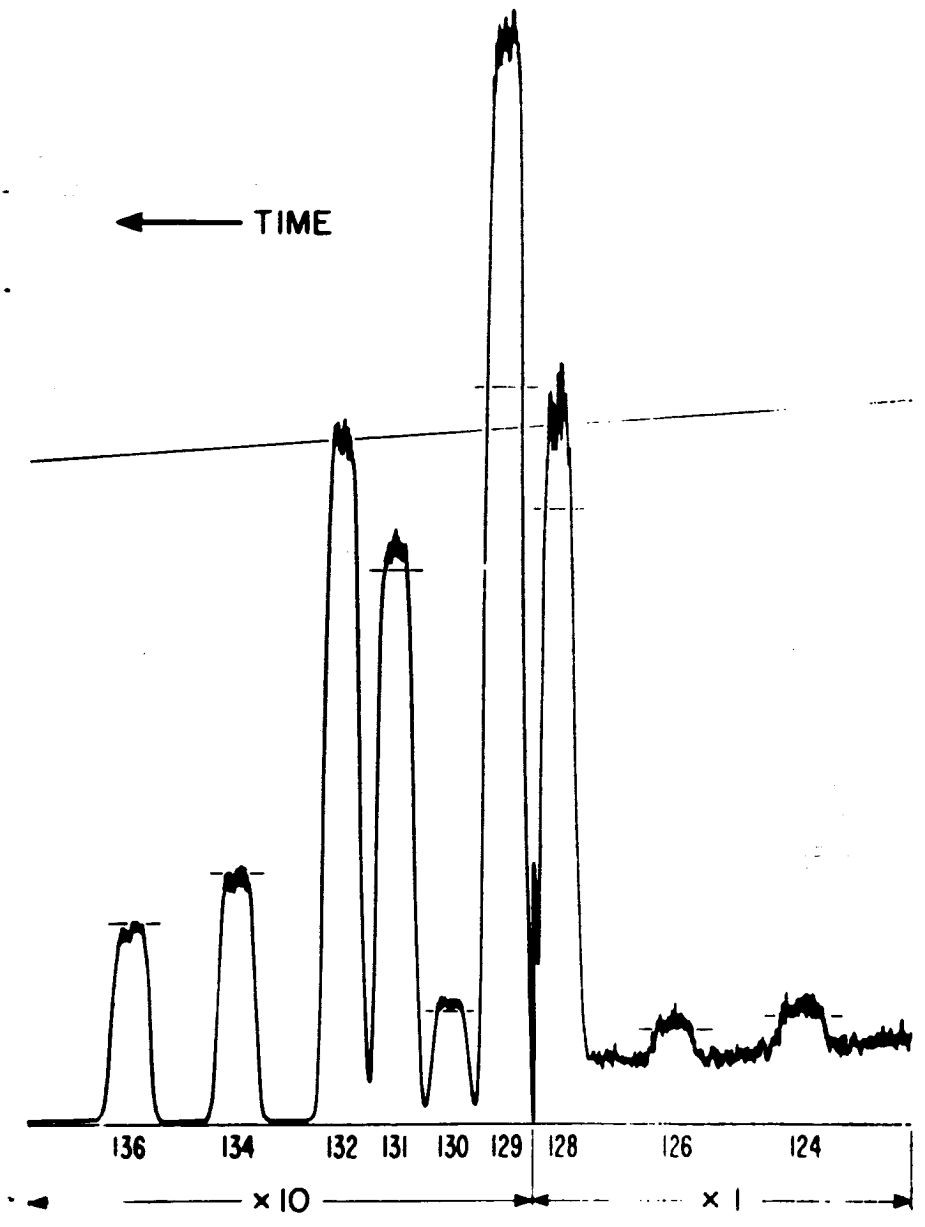


Fig. 13—Mass spectrum of Xe extracted from Richardton stone meteorite. Horizontal lines show the comparison spectrum of terrestrial Xe. I^{129} is enriched by about 50% over terrestrial abundance. From Reynolds, *Physical Review Letters* 4, 6. 8, 1960.

would be I^{129}, with a half-life of 1.72×10^7 years. The presence of this radiogenic Xe^{129} permits another type of date or age to be computed. This is the time interval between nucleosynthesis and the cooling of the parent meteorite planet, cooling to a temperature where Xe^{129} could be retained. It is necessary to make an assumption about the amount of I^{129} created during nucleosynthesis, but this can be done within tolerable limits of error. It turns out that the time interval between nucleosynthesis and planetary cooling was of the order of 10^8 years!

Thus, Anders' hypothesis is reinforced not only by the demonstration that extinct radionuclides other than Al^{26} actually did exist, but because apparently a requirement is set for such rapid heating and cooling that only his hypothetical short-lived radionuclides are capable of filling the bill.

During the past year, discoveries of other extinct radioactivities in meteorites have reinforced the concept. Prominent among these has been the anomaly in isotopic composition of silver reported by Rama Murthy (22), the excess Ag^{107} apparently being the daughter of extinct Pd^{107}. However, recent more detailed work has begun to make doubtful some of the earlier interpretations of these anomalies. Confusing isotope ratios have appeared, deficiencies of certain isotopes where excesses were expected. Apparently it will be some time before the meaning of these radiogenic isotopes is fully understood.

One more body of data bears on the question of thermal history, however. Mention was made earlier that the phase configuration in iron meteorites allows an estimate to be made of the rate at which they cooled. In figure 14 we see a sawed slab of a typical iron meteorite, the Edmonton Octahedrite. The sawed surface has been carefully polished and etched with a nitric acid solution. This procedure brings out the curious array of parallel and intersecting lines known as the Widmanstätten pattern. The narrow parallel lines in Edmonton are plates of kamacite, which is alpha or body-centered cubic nickel-iron. The material intervening between the kamacite plates consists mostly of taenite, gamma or face-centered cubic nickel-iron.

Figure 15 is a photomicrograph of the Dexter iron meteorite. Here we see an element of the Widmanstätten pattern at close quarters. The triangular light-colored area consists of taenite, the darker surrounding areas are kamacite. The darker central portion of the triangular taenite field is plessite, which is a very fine-grained mixture of kamacite and taenite. The nickel-iron phase diagram shows that during cooling of a nickel-iron object both kamacite and taenite phases must change in composition. Kamacite and taenite, in order to maintain thermodynamic equilibrium, must both increase their nickel content as they cool. In order for this to happen the amount of kamacite phase must increase at the expense of the amount of taenite phase. Cooling, therefore, involves the migration of grain boundaries and the diffusion of nickel away from these moving grain boundaries.

Now that electron probe microanalyzers have become available, it is possible to investigate on a microscale compositional variations of the nickel-iron phases in the vicinity of the phase boundaries. In figure 16 we see the result of a traverse run across the surface of the Grant iron meteorite by

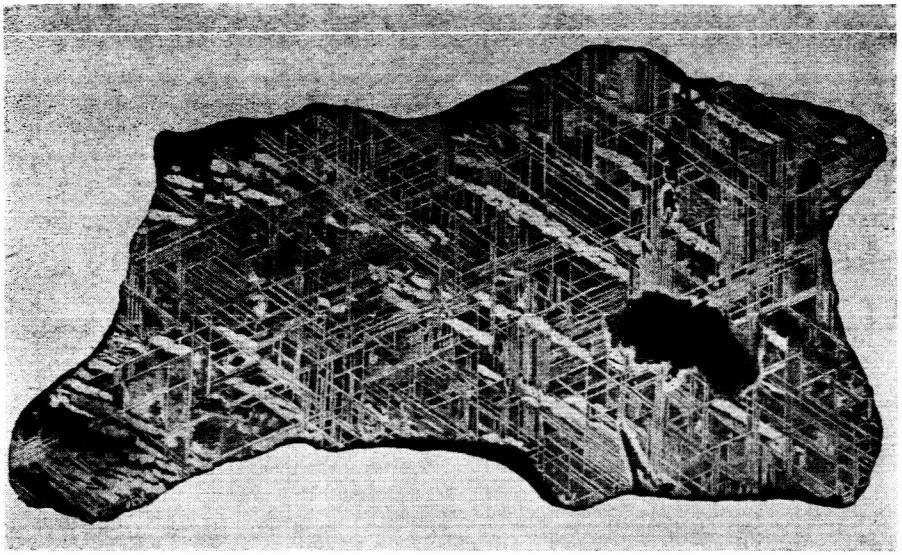


Fig. 14—The Edmonton, Kentucky Octahedrite, showing Widmanstatten patterns. Meteorite is about 20 cm wide. From Uhlig, *Geochimica et Cosmochimica Acta* 6, p. 282, 1954.

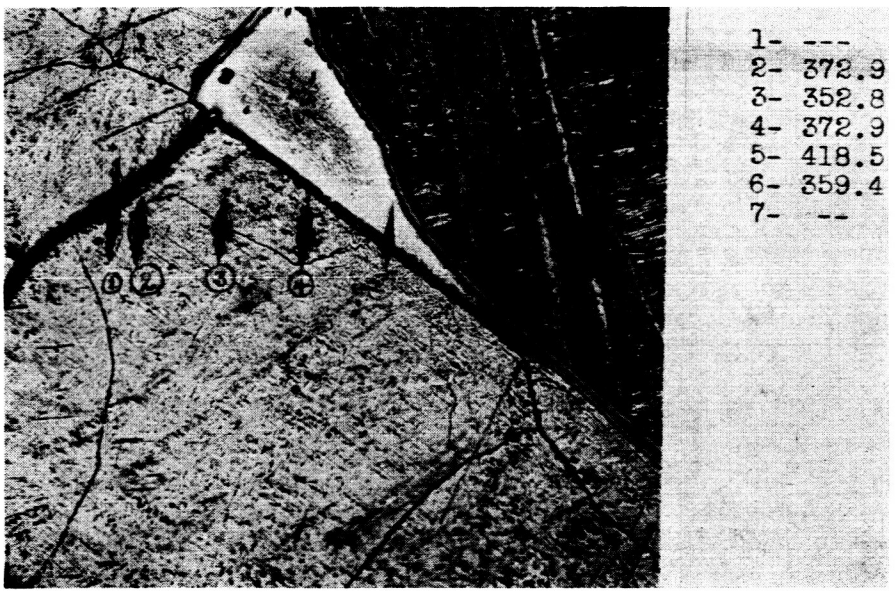


Fig. 15—Dexter (iron) meteorite, in polished section. The field is one mm. wide. The light triangular area is taenite, the surrounding fields kamacite. The darkened center of the taenite area is plessite, a very fine-grained mixture of kamacite and taenite. Numbers identify Knoop hardness impressions. From Uhlig, *Geochimica et Cosmochimica Acta* 6, p. 282, 1954.

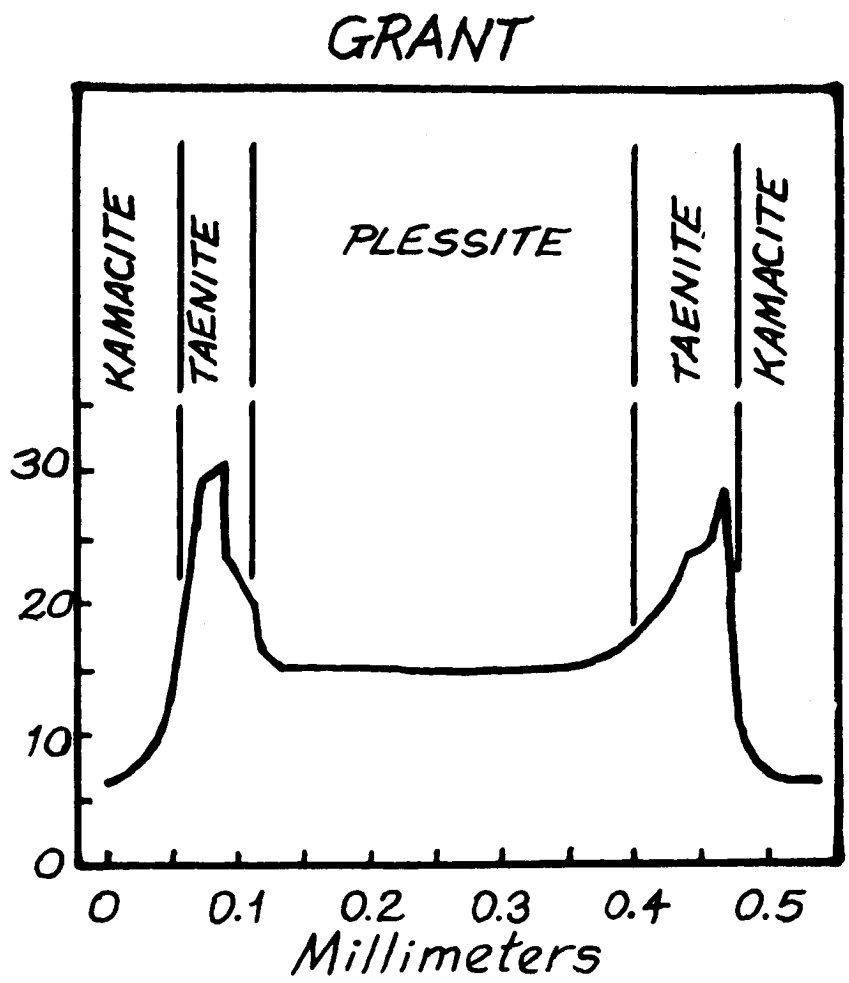


Fig. 16—Variation of nickel content in the kamacite, taenite and plessite zones of the Grant iron meteorite (23), a typical case. Non-uniform composition of the taenite indicates cooling was too rapid for equilibrium conditions to be maintained in that phase.

H. H. Uhlig and M. Feller-Kniepmeyer (23). This traverse has crossed a taenite field which contained a plessite center much like the triangular area seen in figure 15. The electron beam width was not narrow enough to resolve individual kamacite and taenite particles in the plessite region. The interesting thing to note here is that composition within the taenite regions is not constant. Instead, the taenite contains a maximum of nickel where it abuts on kamacite phase, and the amount of nickel decreases away from this phase boundary. This tells us that the cooling rate was too rapid for equilibrium to be maintained within the taenite fields (by means of solid diffusion of nickel atoms).

Since measurements have been made of the diffusion constant of nickel in taenite it should be possible to make an estimate of the cooling rate of nickel-iron meteorites such as Grant. An analytical solution of the diffusion equation would not tell the whole story here. The diffusion equation does not take account of the fact that phase boundaries migrated. This consideration is important, for as the grain boundary migrates it continually destroys the high-nickel portion of the taenite band. In an attempt to obtain as accurate an estimate as possible of the cooling rate of nickel iron meteorites, I have therefore written a digital computer program which yields nickel diffusion curves in taenite phase and which takes account of the migration of grain boundaries. Initial results indicate that the nickel-iron meteorites cooled from its freezing temperature through 1000° C in a period of the order of 10^7 years. This rate of cooling is consistent with the Xe^{129} ages already mentioned, which were of the order of 10^8 years. Thus it is also consistent with Anders' concept of heating by extinct radioactivity.

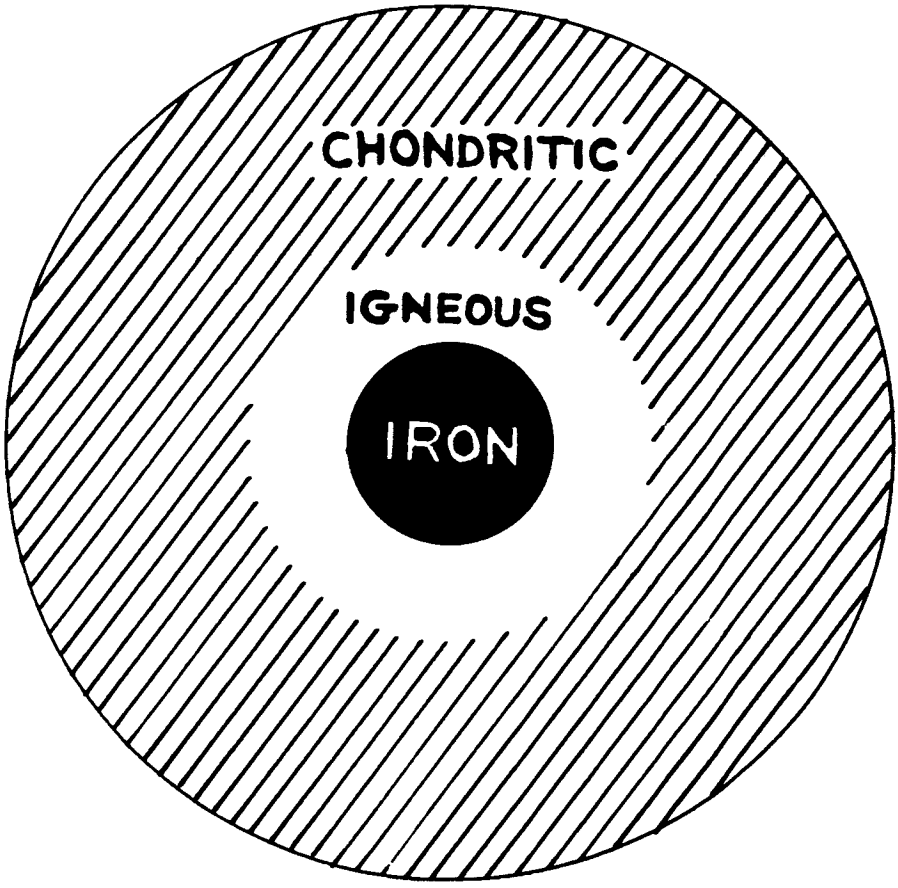


Fig. 17—Sketch of hypothetical parent meteorite planet, with mantle of chondritic rock, and hot core in which molten silicates and nickel-iron have differentiated into two layers.

I mentioned earlier a thermal history which might be considered as an alternative to this mechanism of heating by extinct radioactivity. In this alternative mechanism the parent meteorite planets are heated by the disintegration of U, Th and K^{40}, then disrupted by collision just 4.5 billion years ago and rapidly quenched by radiation of heat into space. It turns out that such a history is improbable but not impossible. If the parent meteorite body were initially 300 to 350 kilometers in radius and if on breaking up it yielded fragments as large as 25 kilometers in radius, then such fragments would be large enough to cool at a rate of about 1000° C per 10^7 years. The trouble is the core of the planet, including the pure nickel-iron zone, must necessarily have been molten at the time of this hypothetical disruption. It is logical to assume that such a molten zone would be spattered into droplets rather than retained in 25 kilometer chunks during disruption.

But let's leave thermal history behind now and go off on another tangent. Figure 17 gives you an idea of the type of parent meteorite planet many of us picture. This has a mantle of chondritic meteorite material and a central zone which has been melted out by heat, whatever its source. This molten central zone has differentiated into two immiscible fluids, the denser liquid iron sinking to form an iron core and the lighter igneous melt forming an intermediate zone between the core and the mantle. The planet must have consisted largely of chondritic material since chondrites are by far the most important meteorite type (Figure 1). Therefore, it would be appropriate to take a closer look at just what these chondrites consist of.

Figure 18 is a photograph of Bjurböle, a typical chondrite. Note that it consists of an agglomeration of small spheres or balls of stone, called chondrules. Hence the name chondrite. Several detached chondrules can be seen next to the chondrite.

Figure 19 is a thin-section of the chondrite Renazzo. The light areas are transparent silicate chondrules composed of the minerals olivine, orthopyroxene, and feldspar. The dark spots scattered within these light areas are small grains of nickel-iron. The black opaque material which intervenes between the chondrules is composed of exceedingly fine grains of the order of 200 Å in diameter. The minerals which make up these grains are not well known, but magnetite (Fe_3O_4) is one of them. Note that one of the chondrules in Figure 19 is armoured with a shell, 0.2 mm thick, of nickel-iron. The thin-section of Tieschitz, Figure 20, shows this chondrite to be similar to Renazzo in that it consists of spherical chondrules embedded in a black opaque matrix. Here the chondrules have a somewhat different appearance from those in Renazzo.

However, thin-sections of the chondrites Lumpkin and Milena, Figures 21 and 22, have a different appearance. Here there is no black opaque intervening matrix; instead, the chondrules are part of a more homogeneous fabric of silicate crystals. The chondrules themselves are no longer clearly delineated, but seem to fade gradually into the rock surrounding them.

Why should there be this difference in textures between chondrites? The answer is probably metamorphism (24). Most meteorite petrographers agree now that many of the chondrites have undergone thermal meta-

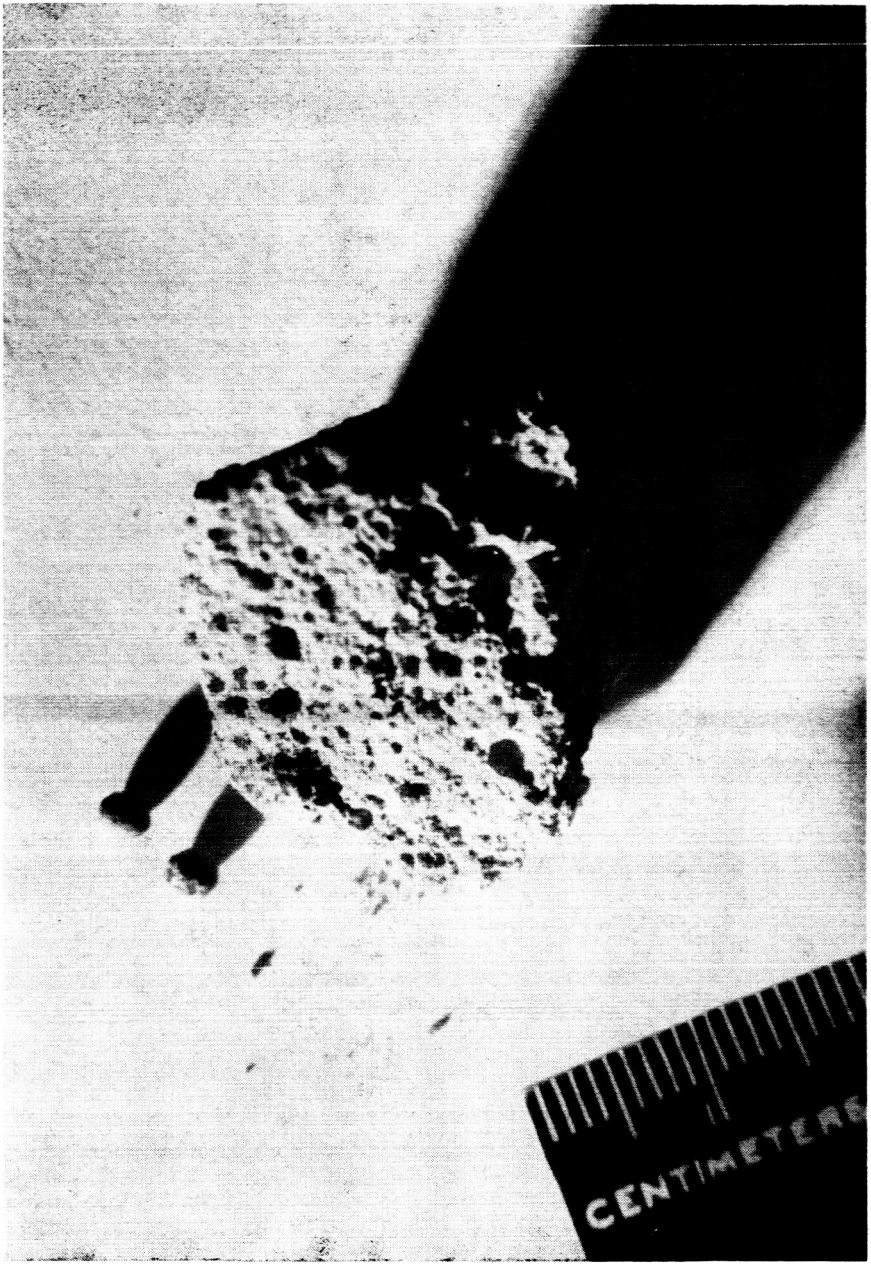


Fig. 18—Bjurböle, a typical chondrite, and several detached chondrules

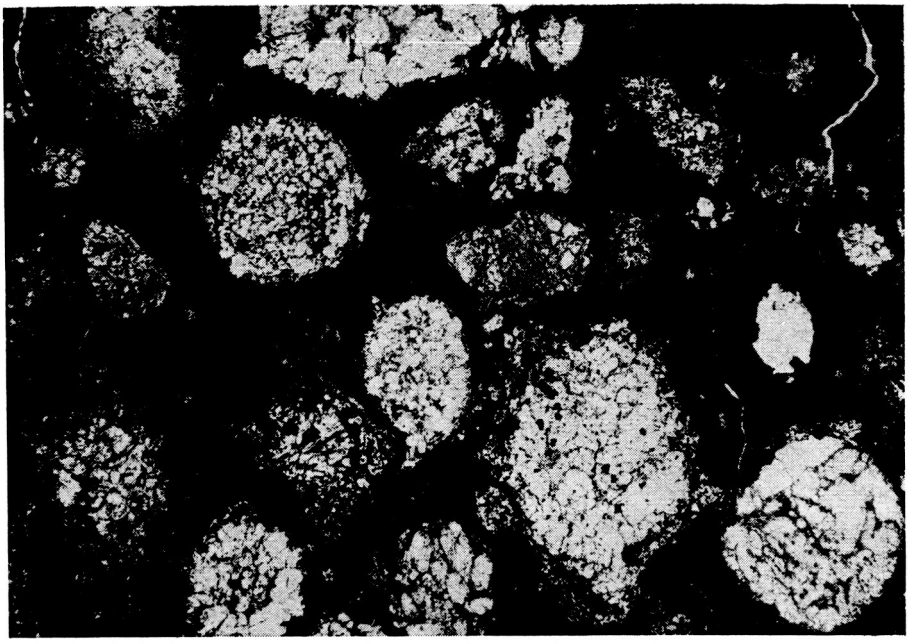


Fig. 19—Thin-section of Renazzo; magnification, 23X. Light areas are (transparent) silicate chondrules, composed of olivine, orthopyroxene, and feldspar. One chondrule is "armored" with a shell (0.2mm) of nickel-iron.

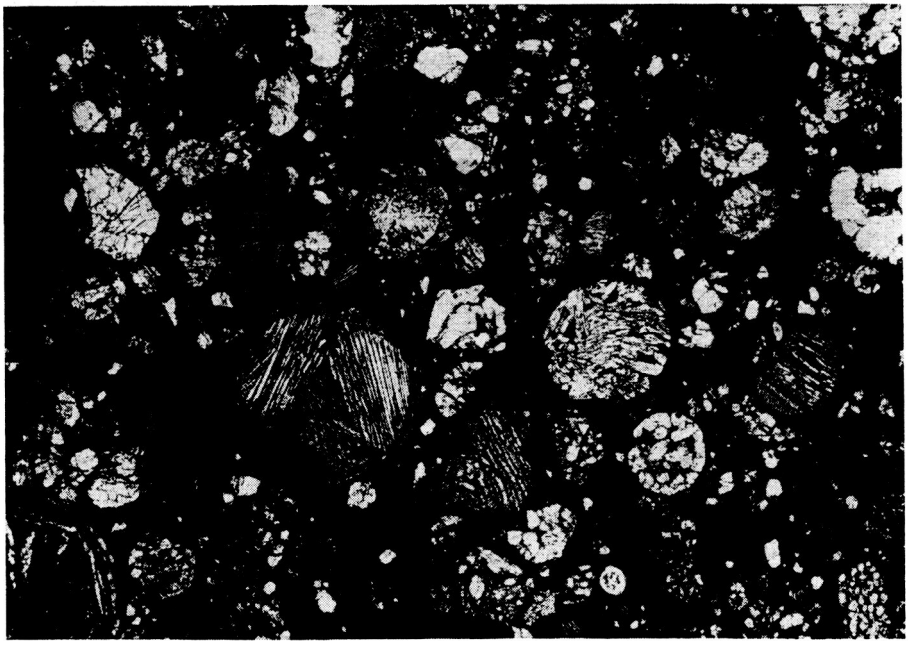


Fig. 20—Thin-section of Tieschitz. Magnification, 23X.

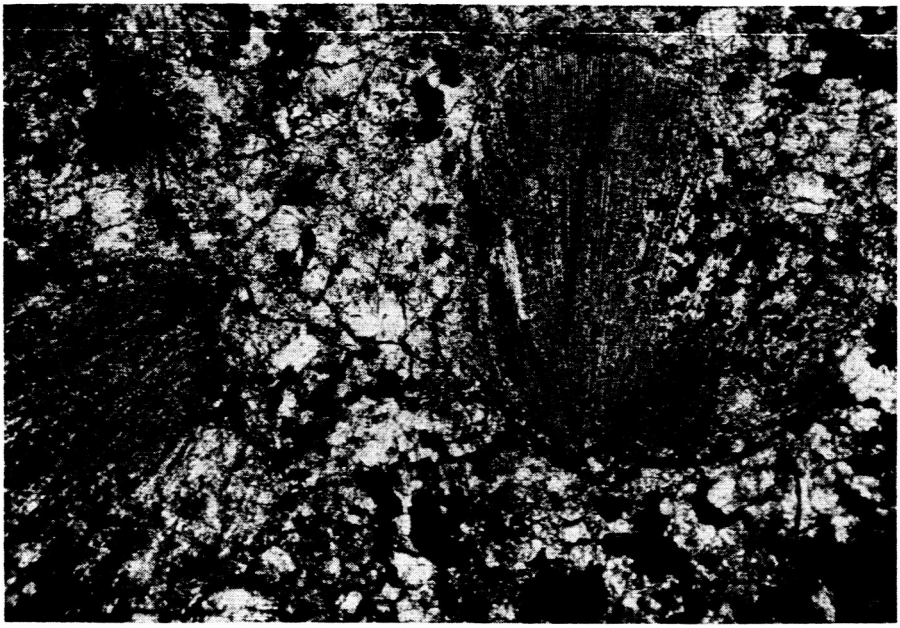


Fig. 21—Thin-section of Lumpkin. Magnification, 45X.

Fig. 22—Thin-section of Milene. Magnification, 45X.

morphism or recrystallization, which tended to make their textures more homogeneous and to erase to some degree the distinctive pattern of chondrules. If this is so, then chondrites such as Renazzo and Tieschitz have been less altered, less tampered with by heat than have other chondrites; therefore they are more representative of the original primordial material of the planets.

Apparently this primordial material consisted largely of chondrules: therefore it is my feeling that the key to the whole problem of the origin of the meteorites, and indeed of the planets and the solar system is to be found in these chondrules (Figure 23). A number of different ideas have been advanced to account for these chondrules, although practically no research has been done on individual chondrules themselves. Here briefly are some of these ideas:

1) That the chondrules formed somehow in situ. The apparently random variation of mineralogy and texture between juxtaposed chondrules in any given chondrite would appear to make such a mechanism improbable. Also, there is no known natural process operating on terrestrial rocks which would be capable of producing the curious chondrule fabric.

2) That they are fragments. Figure 24 casts doubt on this concept. Here we see two chondrules which have become stuck together, indicating that at one time the chondrules were plastic.

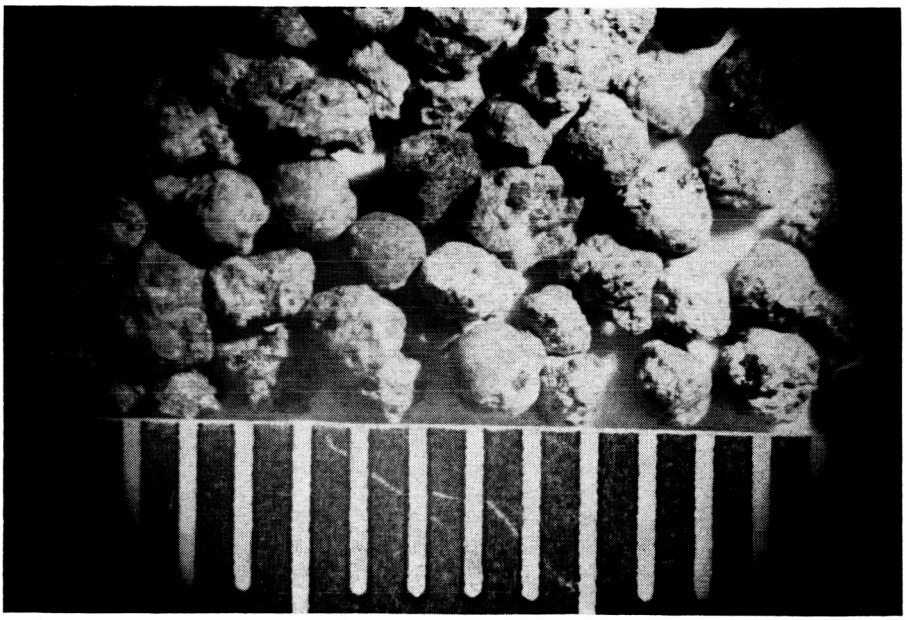


Fig. 23— —20 +30 Sieve fraction of chondrules separated from Bjurböle, with millimeter scale.

Fig. 24—Thin-section of a compound chondrules separated from Bjurböle. Long dimension of the chondrule is about 1 mm. The fibrous mineral grains are orthopyroxene. From Wood, *Geochimica et Cosmochimica Acta* (in press).

3) That the chondrules are pyroclastics; that is, that they are hardened droplets of lava which were thrown out of the throat of a volcano. This concept recognizes the evidence that chondrules were plastic at one time and that they are probably hardened liquid droplets. An argument which I feel can be raised against this mechanism is that the chondrules never contain bubbles, whereas terrestrial pyroclastic fragments are usually to some degree vesiculated.

4) Finally, my own feeling is that these liquid droplets may well have condensed from a hot silicate vapor. All cosmological hypotheses envisage the creation of the elements at very high temperatures, followed by cooling and condensation of the less volatile elements into solid matter, ultimately

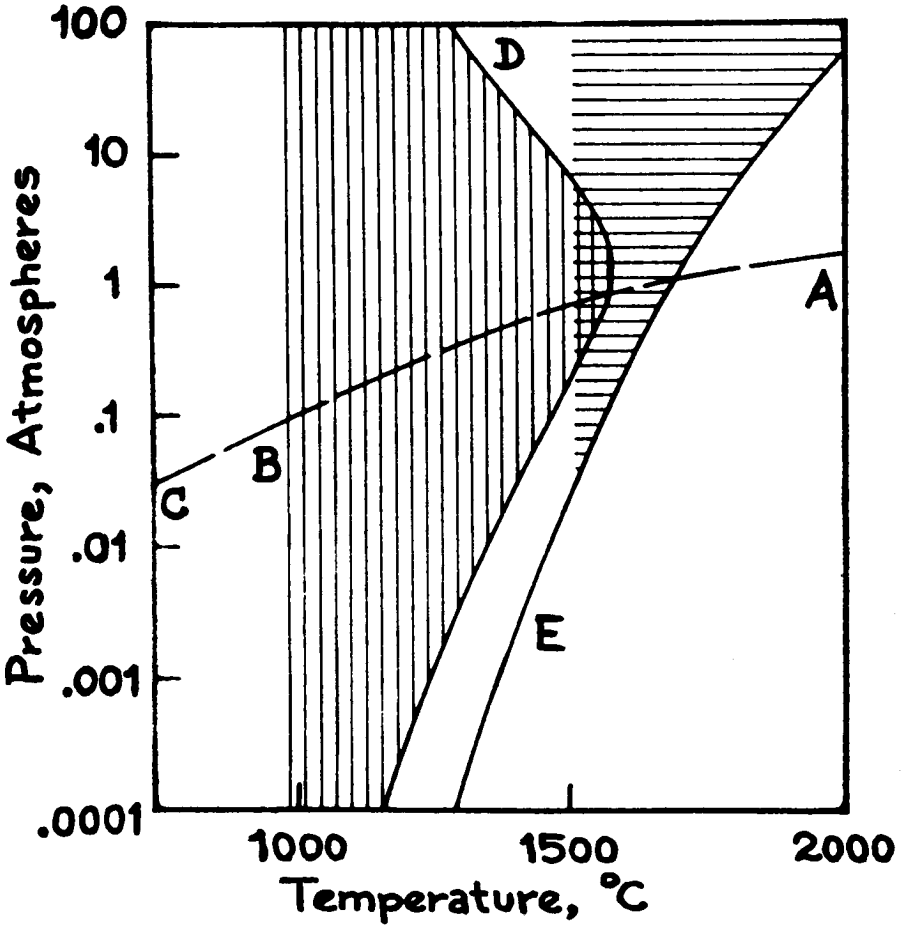


Fig. 25—Pressure-temperature plot showing approximate positions of fields of liquid nickel-iron (horizontal hatching) and liquid silicate (vertical hatching) in an atmosphere of solar composition (27, 28). From Wood, *Geochimica et Cosmochimica Acta* (in press).

planets. If total gas pressure in the system were high enough during cooling, there is a possibility that some condensing primordial matter would pass through a liquid state. Figure 25 is an approximate phase diagram which delineates the liquid field in such a cooling system. In this case, the resemblence of chondrules to hardened liquid droplets, and the preponderance of the chondritic type of meteorite would be explained. Chondrules might be surviving planetesimals, representatives of the first, original, primordial solid matter which accreted to form the terrestrial planets, including the earth. I consider this a very exciting possibility.

References

1. Hawkins, G. S., *Astron. J. 65*, 318, 1960.
2. Goles, G. G., Fish, R. A. and Anders E., *Geochim. et Cosmochim. Acta* 18, 177, 1960.
3. Wood, J. A., *M. N. R. A. S.* 122, 79, 1961.
4. Ceplecha, Z., *Briefl. Mitteilung* 23, Sept. 1959.
5. Kirkwood, D., *The Asteroids*. J. B. Lippincott Co., Philadelphia, 1888.
6. Kuiper, G. P., Fujita, Y., Gehrels, T., Groeneveld, I., Kent, J., van Biesbroeck, G. and van Houten, C. J., *Astrophys. J. Suppl. Series* 3, 289, 1958.
7. Groeneveld, I. and Kuiper, G. P., *Astrophys. J.* 120, 529, 1954.
8. Kuiper, G. P., *Astron. J.* 55, 164, 1950.
9. Urey, H. C., *Astrophys. J.* 124, 623, 1956.
10. Ringwood, A. E., *Geochim. et Cosmochim. Acta* 24, 159, 1961.
11. MacDonald, G. J. F., Harvard University Ph.D. Thesis, 1954.
12. Nininger, H. H., *Arizona's Meteorite Crater*. American Meteorite Museum, Sedona, Arizona, 1957.
13. DeCarli, P. S. and Jamieson, J. C., *Science* 133, 1821, 1961.
14. Fish, R. A., Goles, G. G. and Anders, E., *Astrophys. J.* 132, 243, 1960.
15. Patterson, C., *Geochim. et Cosmochim. Acta* 10, 230, 1956.
16. Schumacher, Z. *Naturf.* 11a, 206, 1956.
17. Geiss, J. and Hess, D. C., *Astrophys. J.* 127, 224, 1958.
18. Fireman, E. L., *Planet. Space Sci.* 1, 66, 1959.
19. Stoenner, R. W., Schaeffer, O. A. and Davis, R., *J. Geophys. Res.* 65, 3025, 1960.
20. Fireman, E. L. and Defelice, J., *J. Geophys. Res.* 65, 3035, 1960.
21. Reynolds, J. H., *Phys. Rev. Letters* 4, 8, 1960.
22. Murthy, V. R., *Phys. Rev. Letters* 5, 539, 1960.
23. Feller-Kniepmeyer, M. and Uhlig, H. H., *Geochim. et Cosmochim. Acta* 21, 257, 1961.
24. Wood, J. A., Smithsonian Inst. Astrophys. Obs. Technical Rpt. 10, 1958.
25. Prior, G. T. and Hey, M. H., *Catalogue of Meteorites*. British Museum of Natural History, London, 1953.
26. Uhlig, H. H., *Geochim. et Cosmochim. Acta* 6, 282, 1954.
27. Brewer, L., Report for the Manhatten Project MDDC–438 C.
28. Urey, H. C., *The Planets*. Yale University Press, New Haven, 1952.

COORDINATES FOR MAPPING THE DISTRIBUTION OF MAGNETICALLY TRAPPED PARTICLES*

CARL E. MCILWAIN
Department of Physics and Astronomy
State University of Iowa
Iowa City, Iowa

Abstract

Dipole representations of the earth's magnetic field have been found to have insufficient accuracy for the study of magnetically trapped particles. A coordinate system consisting of the magnitude of the magnetic field B and the integral invariant I has been found to adequately organize measurements made at different geographic locations. It is shown in the present paper that a parameter $L = f(B,I)$ can be defined which retains most of the desirable properties of I and which has the additional property of organizing measurements along lines of force. Since the parameter L is the analog of a physical distance in a dipole field (the equatorial radius of a magnetic shell), it is usually found to present fewer conceptual difficulties than the integral invariant I.

Introduction

Considerable difficulties were encountered in the early attempts to map the measured intensities of trapped particles. The difficulties were chiefly due to the high spatial gradients in the intensities. At low altitudes in the inner zone for example, a change in the distance from the center of the earth of 3 percent results in an intensity change of over a factor of 10. The shape of the earth's field deviates more than 3 percent from any possible dipole representation, therefore attempts to use a dipole model for a coordinate system can be expected to produce rather chaotic results.

The present paper describes a coordinate system which not only takes account of the nondipole character of the field but can also be used to organize measurements along lines of force as is necessary for most theoretical studies.

Adiabatic Invariants

The use of adiabatic invariants in describing the motion of a particle in a magnetic field has been discussed by Northrup and Teller [1959]. In the present paper, the integral (or longitudinal) invariant at a point A in space

*Supported in part by the Office of Naval Research (contract number N9onr 93803) and by the National Aeronautics and Space Administration (contract number NASw-17).

is defined as $I = \int_{A}^{A'} (1 - B_1/B)^{1/2} ds$, where ds is the differential path length along the line of force connecting the point A with its conjugate point A′, B_1 is the magnitude of the magnetic field along the line of force, B is the magnitude of the magnetic field at the point A, and where the integral is to be taken along the line of force between the conjugate points. Defined in this manner, I can be considered to be a scalar field which has a definite value at each point in space and does not require reference to the motion of trapped particles. If there is no electric field, this definition is equivalent to $I = P^{-1} \int_{A}^{A'} P_{\parallel} ds$, where P is the total momentum of a particle which mirrors at the point A in space and where $P_{\parallel}$ is the component of momentum along the line of force. If we assume that all three adiabatic invariants are conserved for a given particle, then the values of I are the same at every point in space at which a given particle mirrors. If the energy and mirror point distributions of a set of trapped particles in time equilibrium do not change importantly during the time the slowest particles drift once around the earth (for example, the high energy protons in the inner zone), then the directional intensity perpendicular to the line of force will be the same at all points in space which have the same values of B and I.

Magnetic Shells

The points in space which have the same value of B and I form a ring in each hemisphere. A particle mirroring at this B and I will remain upon the surface (or "shell") described by the lines of force which connect these rings.

Now in general, two particles which initially mirror at different values of B along a particular line of force will not drift in longitude to the same lines of force. This means that the shells described by the trajectories of the two particles do not coincide. It will be shown however, that in the earth's magnetic field, this effect is relatively small. One important consequence of this fact is that the omnidirectional intensity as well as the directional intensity is constant along the loci of constant B and I.

Definition of the Magnetic Shell Parameter L

The fact that all particles which drift through a given line of force will remain on approximately the same shell throughout their motion leads immediately to the desirability of finding a method of labelling all points in space with a number which is unique for each shell.

Each locus of constant B and I is a line which lies on a fixed shell. Any parameter which is a function of only B and I will therefore have the correct

longitude dependence. The problem of defining a magnetic shell parameter is thus reduced to finding a function f(B,I) which is constant along lines of force. Such a function could be found empirically by examining the functions $I_i = g_i(B)$ along lines of force calculated with a representation of the earth's field. An empirical fit is not necessary however, because the average of the functions $g_i(B)$ around a shell is well represented by the function obtained for a dipole field.

The integral invariant in a dipole field at a magnetic latitude λ on a line of force which has an equatorial radial distance of R_0 is given by

$$I = 2R_0 \int_0^Y \left\{ 1 - \left[\left(\frac{1+3Y_a^2}{1+3Y^2} \right)^{1/2} \left(\frac{1-Y^2}{1-Y_a^2} \right)^3 \right]^{1/2} \right\} (1+3Y_a^2)^{1/2} \, dY_a \qquad (1)$$

or

$$I = R_0 h_1(\lambda) \qquad (2)$$

TABLE I

I^3 B/M	L^3 B/M	I^3 B/M	L^3 B/M
2.37529×10^{-15}	1.00002	4.39567×10^{-2}	1.53691
1.72697×10^{-12}	1.00016	5.42973×10^{-2}	1.58097
3.70298×10^{-11}	1.00045	6.67358×10^{-2}	1.62782
2.37086×10^{-9}	1.00180	8.16455×10^{-2}	1.67767
2.70390×10^{-8}	1.00406	9.94598×10^{-2}	1.73070
1.52215×10^{-7}	1.00722	1.20681×10^{-1}	1.78716
5.82114×10^{-7}	1.01131	1.45893×10^{-1}	1.84729
1.74352×10^{-6}	1.01632	1.75772×10^{-1}	1.91135
4.41257×10^{-6}	1.02228	2.11103×10^{-1}	1.97966
9.87364×10^{-6}	1.02919	2.52793×10^{-1}	2.05255
2.01130×10^{-5}	1.03707	3.01912×10^{-1}	2.13036
3.80506×10^{-5}	1.04596	3.59672×10^{-1}	2.21352
6.78113×10^{-5}	1.05585	4.27504×10^{-1}	2.30245
1.15049×10^{-4}	1.06680	5.07058×10^{-1}	2.39766
1.87314×10^{-4}	1.07881	6.00263×10^{-1}	2.49967
2.94482×10^{-4}	1.09194	7.09352×10^{-1}	2.60911
4.49250×10^{-4}	1.10620	8.36934×10^{-1}	2.72664
6.67702×10^{-4}	1.12165	9.86054×10^{-1}	2.85300
9.69961×10^{-4}	1.13832	1.16025	2.98905
1.38095×10^{-3}	1.15626	1.36367	3.13571
1.93121×10^{-3}	1.17553	1.60116	3.29403
2.65795×10^{-3}	1.19618	1.87843	3.46520
3.60595×10^{-3}	1.21827	2.20216	3.65056
4.82943×10^{-3}	1.24187	2.58020	3.85161
6.39281×10^{-3}	1.26706	3.02183	4.07005
8.37277×10^{-3}	1.29392	3.53799	4.30783
1.08602×10^{-2}	1.32252	4.14164	4.56716
1.39624×10^{-2}	1.35298	4.84816	4.85057
1.78056×10^{-2}	1.38538	5.67587	5.16095
2.25380×10^{-2}	1.41985	6.64657	5.50164
2.83332×10^{-2}	1.45650	7.78642	5.87648
3.53938×10^{-2}	1.49547		

where $Y = \sin \lambda$. Now in a dipole field

$$B = \frac{M}{R_0^3 \cos^6 \lambda}(1 + 3 \sin^2 \lambda)^{1/2} \tag{3}$$

where M is the dipole magnetic moment. By (3) we see that

$$R_0^3 B/M = h_2(\lambda). \tag{4}$$

Equations (2) and (4) give

$$I^3 B/M = R_0^3 B/M \quad h_1^3(\lambda) = R_0^3 B/M \quad h_3(R_0^3 B/M) = h_4(R_0^3 B/M)$$

or

$$R_0^3 B/M = F(I^3 B/M). \tag{5}$$

TABLE I (cont.)

I^3 B/M	L^3 B/M	I^3 B/M	L^3 B/M
9.12680	6.28994	1.21182×10^3	1.25883×10^2
1.07055×10	6.74720	1.40095×10^3	1.41032×10^2
1.25680×10	7.25432	1.62942×10^3	1.58767×10^2
1.47699×10	7.81840	1.90201×10^3	1.79669×10^2
1.73787×10	8.44781	2.23574×10^3	2.04483×10^2
2.04766×10	9.15245	2.64005×10^3	2.34175×10^2
2.41655×10	9.94413	3.14300×10^3	2.70014×10^2
2.85705×10	1.08369×10	3.76361×10^3	3.13689×10^2
3.38471×10	1.18478×10	4.55081×10^3	3.67477×10^2
4.01899×10	1.29973×10	5.54426×10^3	4.34504×10^2
4.78427×10	1.43104×10	6.83524×10^3	5.19133×10^2
5.71146×10	1.58179×10	8.51087×10^3	6.27578×10^2
6.83985×10	1.75576×10	1.07567×10^4	7.68892×10^2
8.21991×10	1.95767×10	1.37789×10^4	9.56603×10^2
9.91695×10	2.19347×10	1.79974×10^4	1.21153×10^3
1.20162×10^2	2.47069×10	2.39555×10^4	1.56683×10^3
1.46301×10^2	2.79896×10	3.27489×10^4	2.07743×10^3
1.79085×10^2	3.19079×10	4.60402×10^4	2.83877×10^3
2.20532×10^2	3.66255×10	6.73169×10^4	4.02645×10^3
2.45179×10^2	3.93495×10	1.02942×10^5	5.98740×10^3
2.73402×10^2	4.23601×10	1.67841×10^5	9.47159×10^3
3.05069×10^2	4.56967×10	2.96984×10^5	1.63048×10^4
3.41513×10^2	4.94049×10	5.95512×10^5	3.17247×10^4
3.82642×10^2	5.35387×10	1.44453×10^6	7.49153×10^4
4.30241×10^2	5.81615×10	5.00559×10^6	2.51887×10^5
4.84311×10^2	6.33486×10	1.47748×10^7	7.32706×10^5
5.47285×10^2	6.91896×10	4.09266×10^7	2.00752×10^6
6.19343×10^2	7.57921×10	1.91226×10^8	9.28013×10^6
7.03884×10^2	8.32855×10	6.50554×10^8	3.12969×10^7
8.01427×10^2	9.18267×10	5.20779×10^9	2.50188×10^8
9.16824×10^2	1.01607×10^2		
1.05125×10^3	1.12862×10^2		

The magnetic shell parameter L is now defined for a point in the earth's magnetic field by the equation

$$L^3B/M = F(I^3B/M) \tag{6}$$

where I and B are to be calculated for the point with a representation of the earth's field, M is the dipole moment of the earth* and F is the function in equation (5) which is calculated with a dipole field.

A set of values for the function F is given in Table I. For accurate computation of L, the following method can be used:

Let

$$\ln\left(\frac{L^3B}{M} - 1\right) = \sum_{n=0}^{n=6} a_n X^n \tag{7}$$

where $X = \ln (I^3B/M)$. Sets of the coefficients a_n for different ranges of X are given in Table II. This method introduces an error in L of less than 0.3 percent for $-\infty < X < \infty$ and less than 0.03 percent for $X < 10$.

TABLE II

	$X < -16$	$-16 < X < 0$	$0 < X < 8$	$8 < X < 21$	$X > 21$
a_0	0.294	0.62290	0.62291	1.0824	−3.04
a_1	0.330	0.43351	0.43416	0.20395	1.00
a_2	0	1.4495×10^{-2}	1.3680×10^{-2}	5.4145×10^{-2}	0
a_3	0	1.2154×10^{-3}	1.4784×10^{-3}	-9.3218×10^{-4}	0
a_4	0	5.9474×10^{-5}	1.2413×10^{-5}	-5.6831×10^{-5}	0
a_5	0	1.5367×10^{-6}	-8.1278×10^{-6}	2.7879×10^{-6}	0
a_6	0	1.5843×10^{-8}	1.4604×10^{-7}	-3.4751×10^{-8}	0

Significance of L

In general, L should be regarded as a parameter which retains most of the useful properties of I exactly and which is also approximately constant along lines of force. To gain some feeling for the meaning of L, it is useful to remember that in the case of a pure dipole field, a magnetic shell is labelled by an L equal to the shell's equatorial radius.

For most purposes, B and L should be used as the principal spatial coordinates. For studies involving intuition, it is sometimes desirable to use a coordinate system which has a closer resemblance to the actual physical geometry. One system of this kind can be obtained by transforming B and L to polar coordinates using the dipole relations

$$B = \frac{M}{R^3}\left(4 - \frac{3R}{L}\right)^{1/2}, \quad R = L \cos^2 \lambda. \tag{8}$$

This relationship is illustrated in Figure 1.

*The value $M = 8.06 \times 10^{25}$ gauss $cm^3 = 0.311653$ gauss R_e^3 ($R_e = 6371.2$ km) has been found to be satisfactory.

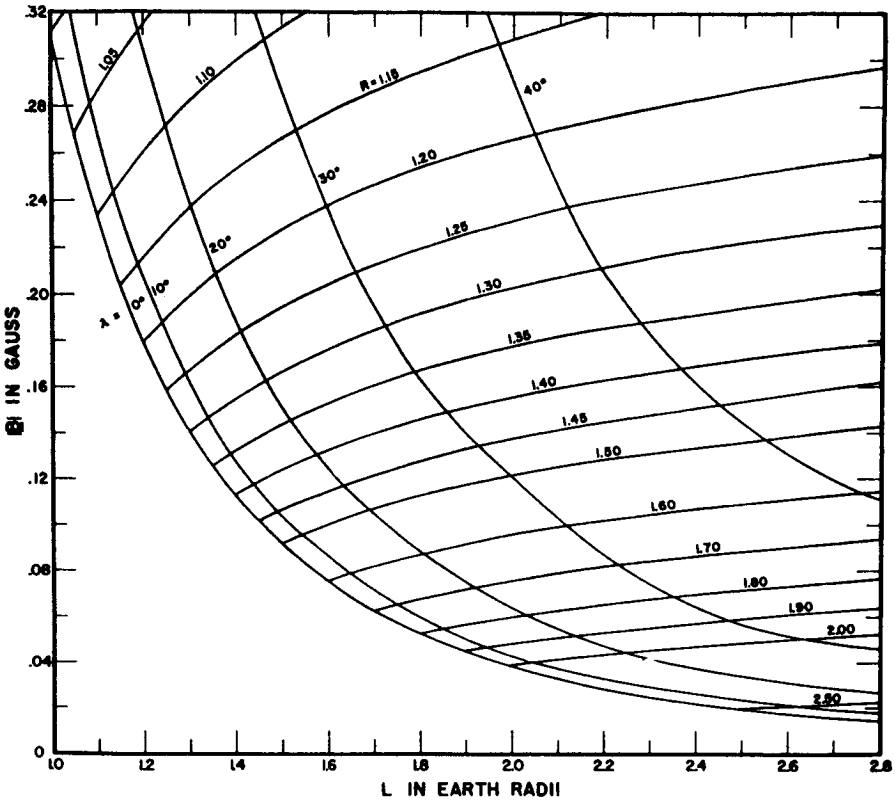


Fig. 1—The mapping of the polar coordinates R and λ on to the B,L plane according to the transformation:
$B = (M/R^3)\ (4 - 3R/L)^{1/2}$, $R = L \cos^2 \lambda$.

A great deal of caution must be exercised when using the coordinates R and λ obtained in this manner, because while the irregular characteristics of the magnetic field will in effect be removed, geographic coordinates will transform in an irregular longitude-dependent manner.

Variation of L Along Lines of Force

Tables of I versus B along 1400 different lines of force were kindly supplied to the author by Messrs. D. C. Jensen, R. W. Murray, and J. A. Welch, Jr. These tables were computed with the 512 term spherical harmonic expansion ($n = 24$, $m = 17$) of the earth's field obtained by D. C. Jensen and W. A. Whitaker [1960] and from the 1955 surface data. The L values obtained from these sets of B and I were found to vary less than one percent along most of the lines of force. Figure 2 shows the variation in L along six lines of force and includes several of the worst cases. The variations along different lines of force were found to be systematic in both

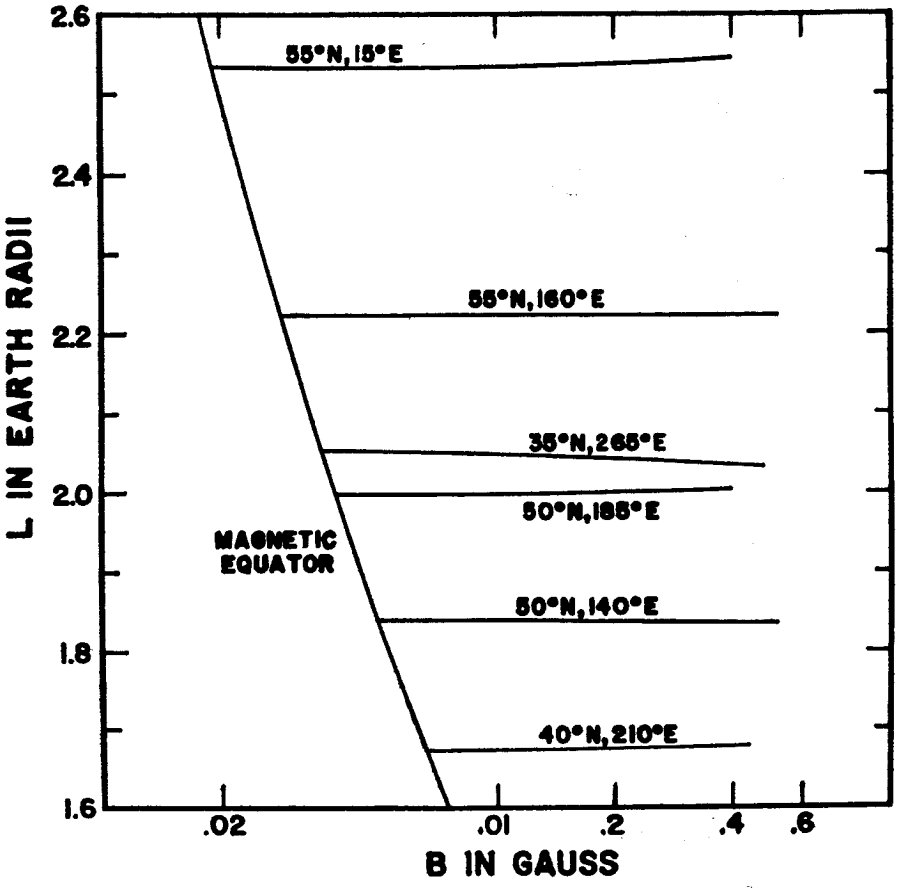


Fig. 2—The variation in L along six lines of force. Several of the worst cases are included.

latitude and longitude. This is illustrated in Figure 3 where the deviations in L, along lines forming the magnetic shells labelled by L = 1.2, 1.5, 2.2 and 3.0 earth radii, are plotted versus the geographic longitude at the northern ends of the lines. A positive deviation means that L increased toward higher values of B.

This kind of check upon the variation of L along lines of force tends to be independent of the precision of the magnetic field analysis. An analysis of the earth's field based upon improved knowledge of the surface field would probably have nondipole terms of similar magnitude and therefore would be expected to give similar variations in L along lines of force. It should be noted however that in the present analysis, the contributions of external current systems have been assumed to be zero.

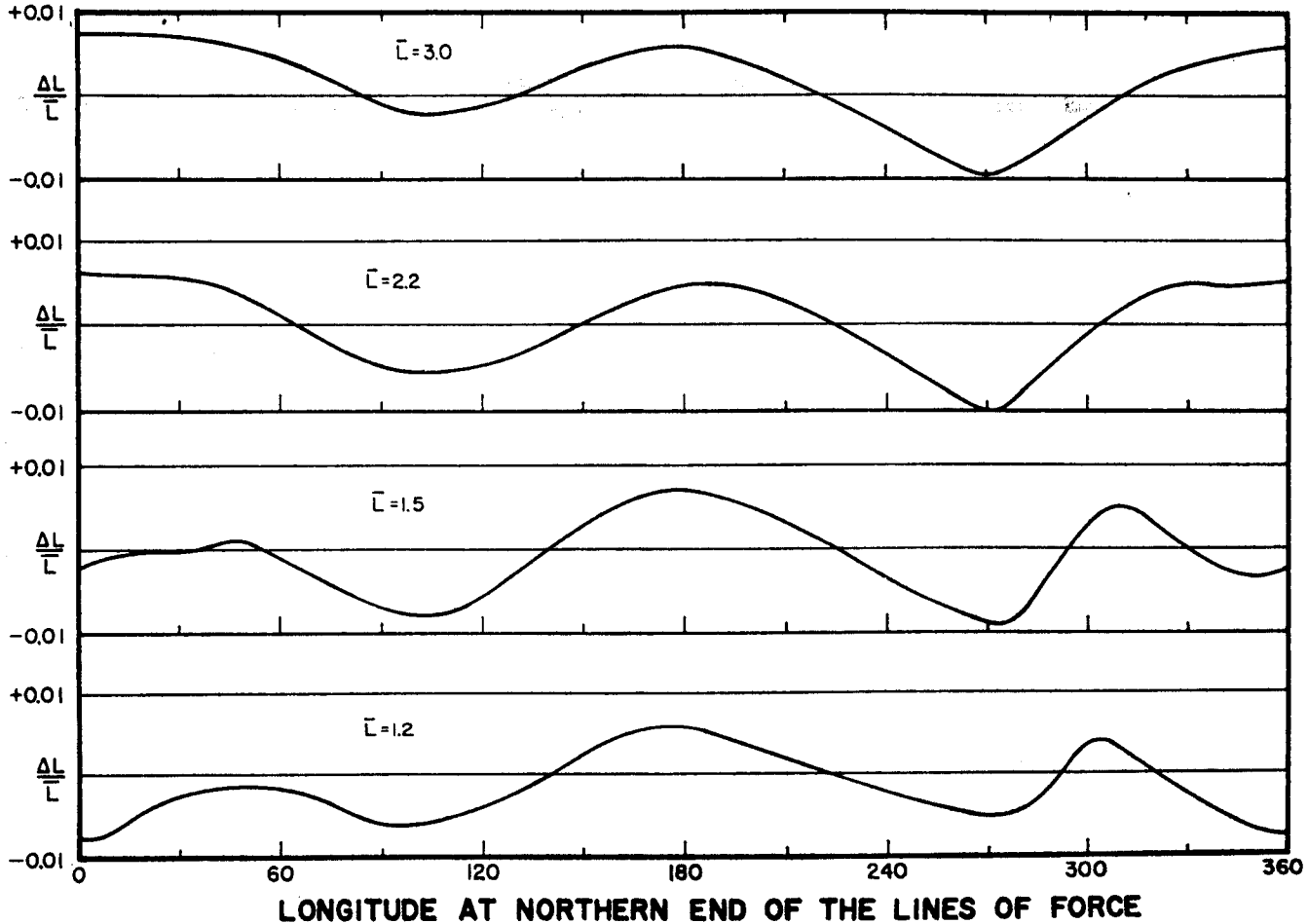


Fig. 3—The maximum deviation in L along lines forming magnetic shells labelled by L = 1.2, 1.5, 2.2 and 3.0 earth radii plotted versus the geographic longitude at the northern ends of the lines. A positive deviation means that L increased toward higher values of B.

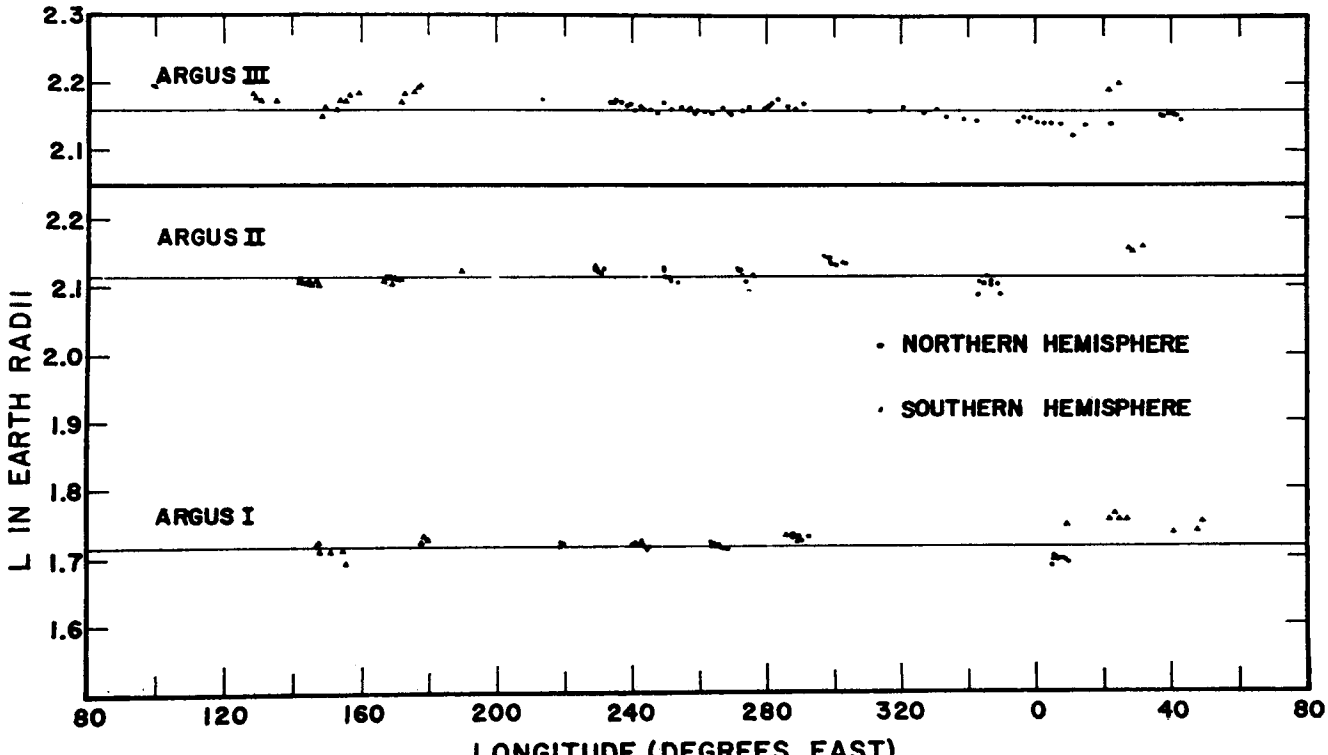


Fig. 4—The L values for the observed intersections of the Explorer IV trajectory with the three Argus shells are plotted versus geographic longitude.

The Argus Shells

Well-defined shells of electrons injected by the three high altitude nuclear detonations Argus I, II and III were detected by Satellite 1958ϵ (Explorer IV) [Van Allen, et al., 1959]. Each of the observed intersections of the satellite trajectory with an Argus shell should have the same value of L. Figure 4 shows the L values obtained for the observed intersections as a function of geographic longitude. The average L values for the shells are 1.715, 2.115 and 2.16 earth radii for Argus I, II and III respectively. In addition to the scatter due to measurement inaccuracies, systematic variations with longitude of approximately $\pm$ 1.5 percent can be discerned. By comparison, Pennington [1961] has found that the trace of the data points in the northern hemisphere for Argus III projected upon the surface of the earth deviates more than $\pm$ 3.5 degrees from the earth trace of the best possible offset dipole shell. A latitude error of $\pm$ 3.5 degrees for the Argus III shell corresponds to an error in L of about $\pm$ 13 percent.

The range of B for the data points on each shell is not large. The deviations of the L values in Figure 4 are therefore almost entirely due to inaccuracies in the surface field upon which the magnetic field representation is based. It is interesting to note that the deviations in the L values are largest in the regions where the discrepancies between the various available maps of the surface field are also large. No definite time dependence can be discerned in the data shown in Figure 4. An upper limit on the rate of change in L is 0.001 earth radii per day.

Proton Intensities in the Inner Zone

The most complete set of data pertaining to the spatial distribution of particles in the inner zone presently available was obtained by the detectors on Satellite 1958 Epsilon [Van Allen, McIlwain and Ludwig, 1958]. The results of a preliminary analysis of the Geiger tube data are shown in Figures 5 and 6. The geometric factors and proton thresholds are approximately 0.54 cm^2 and 31 Mev for the unshielded counter and 0.62 cm^2 and 43 Mev for the shielded counter. Comparison with the scintillation counter data indicates that most of the Geiger tube counting rates were probably due to penetrating protons rather than bremsstrahlung from the very much higher fluxes of non-penetrating electrons.

The data points shown in Figures 5 and 6 were obtained from recordings made at stations in North and South America, Europe, Africa, Asia, Australia, New Zealand and Hawaii. Data obtained during the periods immediately following high altitude nuclear detonations have been excluded.

The general physical configuration of the lower part of the inner zone is illustrated in Figures 7 and 8 where the B versus L curves in Figures 5 and 6 have been transformed into polar coordinates in the manner previously described. The dashed lines approximately represent the maximum excursions of the earth's surface in these coordinates.

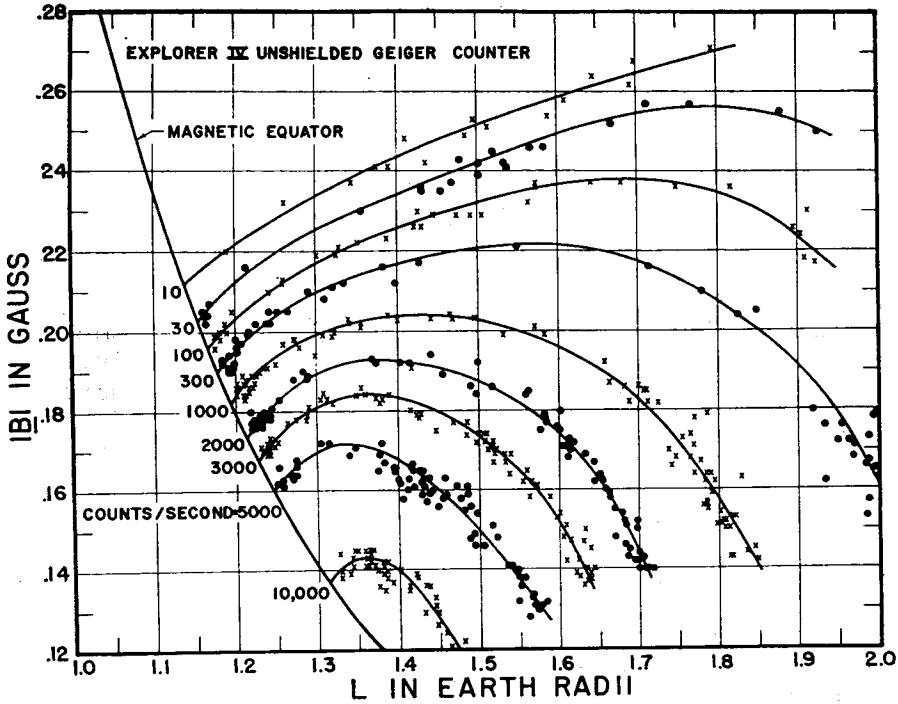


Fig. 5—Contours of constant true counting rate of the unshielded Geiger counter in Explorer IV. The points shown correspond to data obtained over a wide range of geographic latitude and longitude.

Some of the aspects of these graphs which will be subjects for future papers include:

(a) Determination of angular distributions in the manner outlined by E. C. Ray [1960].
(b) The spatial dependence of proton injection and loss mechanisms.
(c) The spatial dependence of the proton energy spectrum.
(d) The time dependence of the intensities.
(e) Comparison with data obtained by instruments on other space vehicles.

Solar Cosmic Rays

The application of the B and L coordinate system is by no means limited to trapped particles. In general, it can be usefully employed in the study of any phenomenon in which motion tends to be along lines of force. Some obvious examples are auroral particles, whistler node wave propagation, and low energy cosmic rays.

The latitude dependence of solar cosmic rays is of particular interest in that it can be used to obtain information about both the energy spectrum of

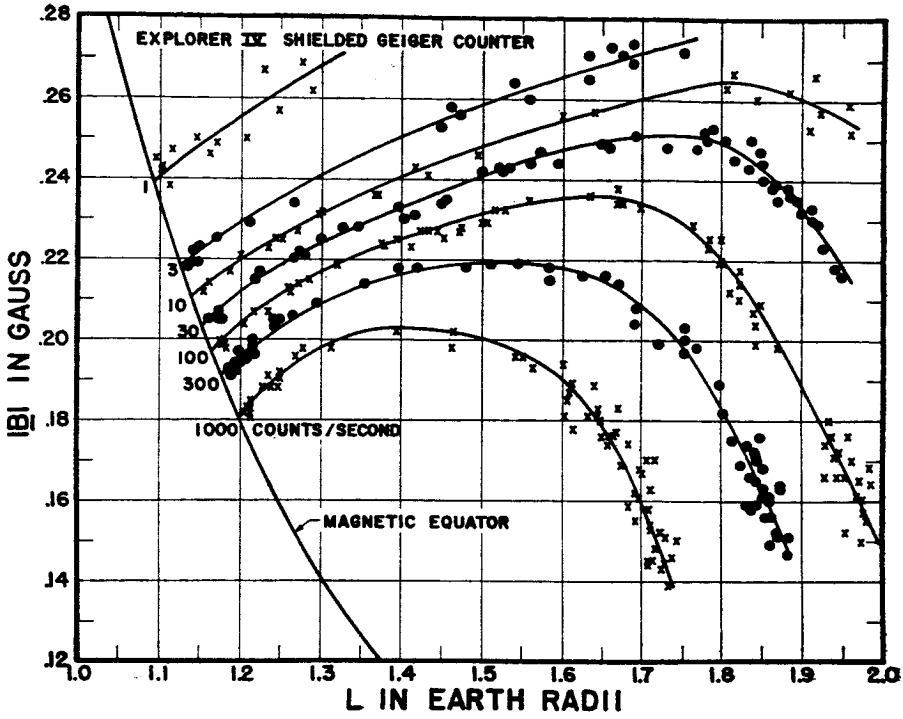


Fig. 6—Contours of constant true counting rate of the shielded Geiger counter in Explorer IV.

the particles and the distortion of the earth's magnetic field by external current systems. An example of the latitude dependence of solar cosmic rays is shown in Figure 9. The left side of this figure shows the intensity versus time measured by the shielded Geiger counter aboard Satellite 1959 Iota (Explorer VII) during a solar cosmic ray event and the right side shows the same data points plotted versus L.

Time Varying Magnetic Field

The magnetic shells labelled by L values greater than 3 earth radii are probably distorted by an important amount during magnetic storms. It would be difficult if not impossible to calculate true instantaneous values of I during a magnetic storm. It is suggested that the values calculated for magnetically quiet periods be used as a fixed reference system in which the perturbations on the various phenomena can be studied.

Computation of I

Direct evaluation of $I = \int (1 - B_1/B)^{1/2}\, ds$ entails a considerable amount of computation. A number of relatively fast methods of computing I have been devised. At least one of these methods (which in general involve various

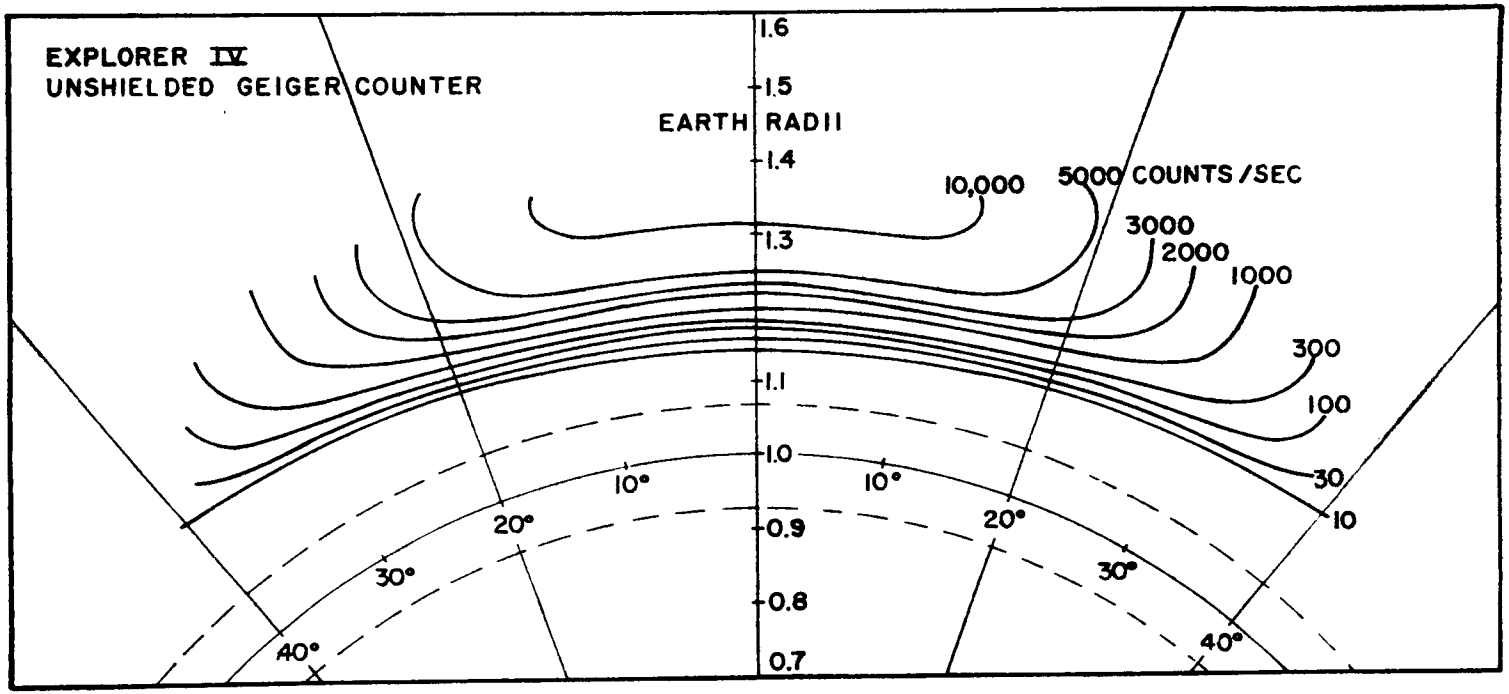


Fig. 7—The contours of constant counting rate shown in Figure 5 transformed to the polar coordinates R and λ.

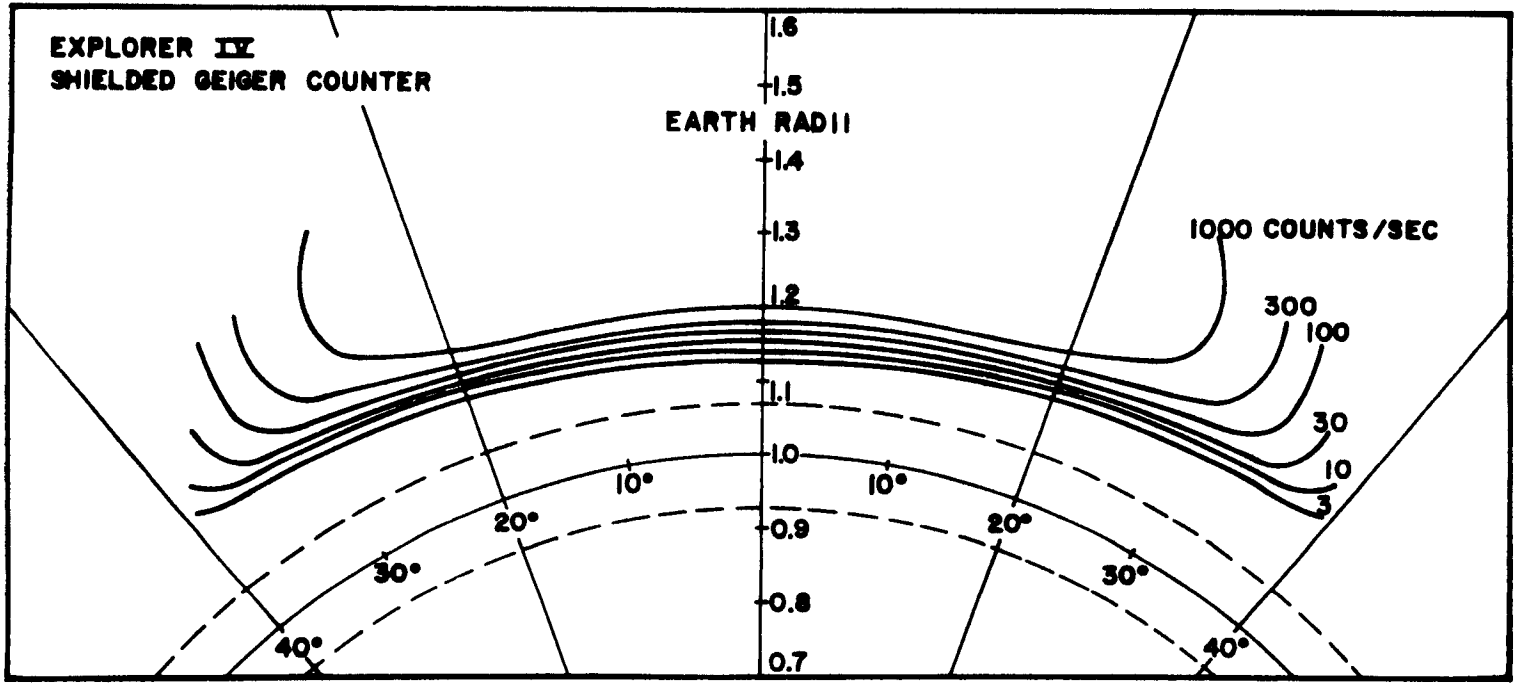


Fig. 8—The contours of constant counting rate shown in Figure 6 transformed to the polar coordinates R and λ.

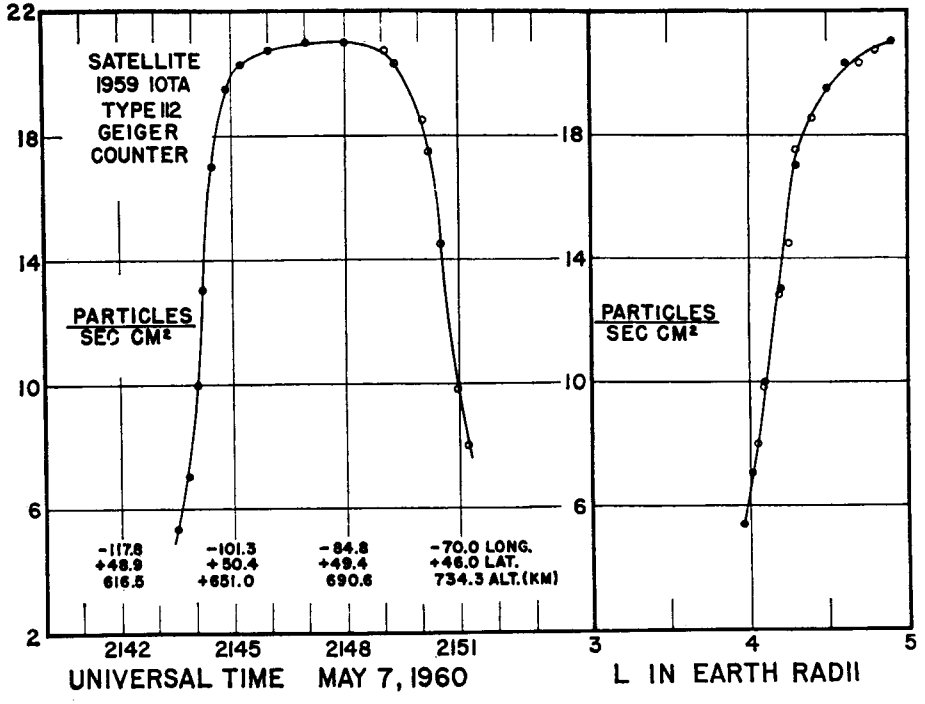


Fig. 9—The left side of this figure shows the omnidirectional intensity measured by the shielded Geiger counter aboard Explorer VII as a function of time during a solar cosmic ray event. The right side of the figure shows the same data points plotted versus L. On some occasions the modification of the earth's field by external current systems permits solar particles to arrive along lines of force labelled by considerably lower L values.

means of interpolation) will probably be made available in the form of a computer program written in the Fortran language.

Conclusion

The choice of B and L as the primary coordinates is of course quite arbitrary. One possible alternate for B is $B/B_0 = L^3B/M$ where $B_0 = M/L^3$ is the equatorial value of B on the line of force. It should be noted that any pair of the quantities B, L, R, λ and B/B_0 can be used in place of B and L. To avoid confusion and to facilitate the comparison of different sets of data, it is suggested that a pair of coordinates other than B and L be employed only if a substantially improved presentation is obtained.

Acknowledgements

I would like to thank Professor J. A. Van Allen and Professor E. C. Ray for their suggestions and helpful discussions. I would also like to thank

Captain J. A. Welch, Jr., Mr. D. C. Jensen and others at the Air Force Special Weapons Center for furnishing many of the necessary computations.

This research is supported in part by the Office of Naval Research (contract number N9onr 93803) and by the National Aeronautics and Space Administration (contract number NASw-17).

References

Jensen, D. C. and W. A. Whitaker, "A Spherical Harmonic Analysis of the Geomagnetic Field", J. Geophys. Research *65*, 2500 (1960).

Northrup, J. G. and E. Teller, "Stability of the Adiabatic Motion of Charged Particles in the Earth's Magnetic Field", Phys. Rev. *117*, 215-225 (1960).

Pennington, R. H., "Equation of a Charged Particle Shell in a Perturbed Dipole Field", J. Geophys. Research *66*, 709-712 (1961).

Ray, E. C., "On the Theory of Protons Trapped in the Earth's Magnetic Field", J. Geophys. Research *65*, 1125-1134 (1960).

Van Allen, J. A., C. E. McIlwain and G. H. Ludwig, "Radiation Observations with Satellite 1958ε", J. Geophys. Research *64*, 271-286 (1959-a).

Van Allen, J. A., C. E. McIlwain and G. H. Ludwig, "Satellite Observations of Electrons Artificially Injected into the Geomagnetic Field", J. Geophys. Research *64*, 877-891 (1959-b).

PHYSICS OF THE PLANETS*

WILLIAM M. SINTON
Lowell Observatory
Flagstaff, Arizona

Introduction

We are now on the eve of man's exploration of the planets by direct or active means. It is the time at which we must assess what we now know about them in order to carry out the exploration in an intelligent and efficient manner. But ground based observation will yet be a useful means of study for a number of years. The inner planets will receive attention first and the outer planets will be studied only later. Rockets are expensive and many things can be learned more cheaply and easily from the ground. We will also need as much prior knowledge as possible in order to best plan experiments to be made with rockets.

On the other hand, rockets offer the only possibility of "in-situ" measurements of such important atmospheric parameters as temperature, pressure, composition, and of magnetic fields. Undoubtedly we will be in for many surprises when such measurements are made, but we should prepare ourselves beforehand with the highest knowledge now obtainable about the planets.

Instrumented rockets will be sent first, but man will surely travel to the planets. To say that scientific instruments can learn everything that a manned rocket can is sheer folly. The design of an experiment is based on an expected model, but an immediate glance by an observer might tell that the model is ridiculous and that some other experiment should be made. If syphilis was really introduced to Europe from the New World, would physicians have learned of it if Queen Isabella had sent remote controlled ships with bacterial probes instead of Columbus and his men?

The exploration of the planets will eventually help with one of the major deficiencies of astronomical knowledge. Decades ago astronomers largely gave up the study of the solar system because they felt that they had sufficient knowledge of the planets and because the physical study of the stars offered many new fields. Now, however, astronomers are wondering more about questions of the origin and evolution of stars. Since the retinue of planets that our sun has offers clues to the solution of this problem, the study of the solar system is once more of interest. A substantial fraction of the stars undoubtedly have similar retinues and the formation of these must play a role in stellar evolution.

---

*Portions of the material in this paper will appear in a chapter on Planetary Surfaces written by W. N. Sinton for THE MOON–ITS ASTRONOMY AND PHYSICS edited by Zdenek Kopal, Academic Press, Inc., New York, 1961. Permission to use this material here has been granted by the Publisher.

TABLE I

Physical Data of Planets and the Moon

Body	Rotation Period	Mass	Radius	Density	Albedo	Gravity	Maximum Temp.	Effective Temp.	$\frac{T_m}{\pi^{1/4}}$
Mercury	88d	0.05	0.38	5.46	0.058	0.38	625	...	...
Venus	?	0.814	0.961	5.06	0.73	0.88	324	229	243
Earth	23h56m	1.000	1.000	5.52	0.39	1.00	349	246	...
Moon	27d 7h	0.012	0.273	3.33	0.072	0.16	387	274	...
Mars	24h37m	0.1069	0.523	4.12	0.148	0.39	307	217	...
Jupiter	9h55m	318.35	10.97	1.35	0.51	2.65	145	102	109
Saturn	10h14m	95.3	9.03	0.71	0.50	1.17	107	76	80
Uranus	10.7h	14.54	3.72	1.56	0.66	1.05	69	49	52
Neptune	15.8h	17.2	3.38	2.47	0.62	1.23	56	40	42
Pluto	6.4d	0.033?	0.45	2.?	0.16	0.16?	60	42	..

We might begin the study of planetary and lunar physics by assembling the general characteristics of the planets and some satellites in Table I. These are data that have been determined by astronomers for a long time, but they serve to set the ground work for a physical study of the planets. The columns give in order, period of rotation, mass, mean radius, density, the albedo, and the surface gravity.

We may use the data in this table in different ways. Estimates of the temperatures of different planets may be obtained. For a planet which maintains the same face to the sun or rotates only slowly the maximum temperature is $T_M = 394°(1 - A)^{1/4}a^{1/2}$. If on the other hand the heat received from the sun is spread uniformly over the surface of a spherical planet the temperature is $T_e = (1/2)^{1/2}T_M$. Another model that we may use is one in which there is rapid rotation giving uniform temperatures distribution between day and night, but in which interzonal circulation is not permitted. The temperature in different latitudes is then $T_\theta = (\cos\vartheta/\pi)^{1/4}T_M$.

The application of these models to different planets is certainly only of approximate value. The assumption of blackbody emission is not correct for most planets with atmospheres and the "greenhouse effect" surely plays an important role. The "effective" temperature T_e does, however, specify the total emission to be expected from a planet when integrated over wavelength and space. In this connection comparison of observed temperatures with the effective temperature will reveal the character of the planetary emission in unobserved regions of the planet's spectrum in order that the balance of heat be maintained.

In connection with the temperatures we can make an estimate of the types of atmospheres expected on different planets. Jeans was the first to do this. His theory has been much improved since by Jones, Spitzer, and others, but, since we will only use this theory as a guide, we will not go into the more refined calculation. The velocity of escape from a planet is given by $V_{esc} = (2GM/R)^{1/2} = (M/R)^{1/2} \times 11.3$ km/sec if M and R are given in earth units. Jeans (1916) found that an atmosphere will be stable for astro-

nomical times if the r.m.s. velocity of the molecules V_m is less than $0.2V_{esc}$. The molecular velocity is given by $V_m = (3RT/\bar{m})^{1/2}$, where R is the gas constant, T is the temperature of the escape layer (where there is only one mean-free path to escape from the planet's gravitation) and $\bar{m}$ is the molecular weight of the gas in question. The temperature of the escape layer will not be directly equivalent to any of the previous temperatures, but it will, however, vary as $a^{1/2}$.

In Table II is a partial list of atomic abundances. It is now generally thought that the planetary system and the sun evolved from a cloud of dust and gas, which initially had the cosmic abundances given in Table II. The massive major planets retained a much larger fraction of hydrogen and helium than did the terrestrial ones (Mercury, Venus, earth, Mars, and Pluto). Loss of hydrogen converts a highly reduced atmosphere to an oxidized one. The initial escape of hydrogen is responsible for the differences between the atmospheres of the major planets (containing H, He, CH_4, and NH_3) and those of the terrestrial ones (containing O_2, N_2, CO_2, CO, etc.). The changes in atmospheres going on at the present can largely be understood in terms of modifications to Jeans' theory. Terrestrial planets that have atmospheres are still losing hydrogen and helium. This can only be understood in terms of slow diffusion to the escape layer and this rate of diffusion now minimizes the loss of these gases. In the case of the moon, still different considerations apply and we shall discuss these later.

One might continue in the present vein and try to predict completely from theory the present structure of the planets. This would no doubt lead us far afield for the present purposes, and I think it better to switch now to the observational approach to the problem.

In the passive astronomical study of the planets (from the light which they send us as contrasted to the active study where radiation or matter is sent to them), one makes use of the known properties of light. The early

TABLE II

Cosmic Abundances of Elements (Kuiper, 1952)

Z	Elem.	log Abundance
1	H	10.10
2	He	9.24
6	C	6.14
7	N	6.48
8	O	6.70
9	F	4.6
10	Ne	6.77
11	Na	4.25
12	Mg	5.59
13	Al	4.50
14	Si	5.56
16	S	5.19
17	Cl	5.1
18	A	5.9

study of the planets was dominated by one property of light—the ability to form images. From these images, usually sketches by the observers, much has been learned of the planets, but the overinterpretation of these drawings led to controversies which could not be settled and led eventually to a disinterest on the part of astronomers in planetary studies. Photography has to a large extent confirmed the detail sketched in these drawings.

The physical study of the planets began essentially in the 1920's with the utilization of photometric, polarimetric, and spectroscopic studies of planetary light and the extension of studies into the infrared and ultraviolet. These methods were often developed first for stellar observations and their application to planetary problems has been sporadic.

Accurate photometry of planets dates back to Zöllner (1865) with later advances by Müller (1897), and E. S. King (1919). These measurements were visual and photographic. Often rather large systematic errors were involved. Photoelectric cells were applied by Stebbins and later by Whitford, and photometry of planets was made by Rougier (1937). After World War II the availability of photomultiplers greatly increased the accuracy and flexibility of photometry, but application of these to planetary work has been slow. For example, Mars has been only infrequently observed and only in recent oppositions, namely 1954, 1956, 1958, and 1960 and largely at Lowell Observatory. Venus has also been observed photoelectrically at Lowell. These measurements are now being prepared for publication. The albedo given for Venus in Table I is an estimated albedo for solar radiation, but the values we have obtained for the U, B, and V, (ultraviolet, blue, and visible) are 0.53, 0.78, and 0.87 respectively. The visual albedo of 0.87 is much higher than that of Müller (0.59) and Danjon (0.72). We believe that there is less likelihood of systematic errors in the photoelectric work and these values serve to illustrate the difficulties. The cases of Venus and Mercury are, however, the worst in this respect because the measurements, in large part, have to be made in daylight. For the heat budget problem the near infrared is of importance as great as the visible, and very little planetary photometry has been done in this region. Of the sun's light 50% lies beyond 7000 A. A difficulty arises for objects other than the moon, Mercury, and Venus. The planets beyond the earth's orbit can not be observed at all phase angles, a necessary condition to determining the albedo without very uncertain extrapolations.

Extensive polarimetry of the moon and planets was performed by Bernard Lyot (1929), and of the moon by F. E. Wright (1927). Lyot succeeded in constructing a very sensitive visual polarimeter, which measured polarizations to an accuracy near 0.1%. His work is now being carried forward by his pupil, A. Dollfus. Some of the individual measures will be discussed later but it is sufficient to say here that the plane of polarization (as defined in U. S. practice as that of the electric field) is generally perpendicular to the plane of the earth, the sun, and the planet. The amount of polarization is dependent on position of the disk and the phase angle (the angle earth, planet, sun). For very small phase angles the plane of polarization is often found to lie in the plane of these objects rather than perpendicular to it, and

the polarization is then termed negative. At angles larger than 20 to 30 degrees, the polarization is usually positive.

As far as I know, the Faraday rotation of the plane of polarization due to a planet's magnetic field has not been detected. It is small, about 5′ for the earth for observations near the zenith and 45′ at a point on the magnetic meridian 5° from the horizon. The rotation effect does offer the possibility of detecting planetary magnetic fields and should be looked for.

Photographic spectroscopic studies have been made with equipment largely developed for stellar work. V. M. Slipher (1929) obtained spectra of the major planets showing immense amounts of methane and ammonia. Studies were also carried out to determine the presence of water and oxygen on Mars and Venus. These later gave negative results with the possible exception of H_2O on Mars. He also determined the rotation of Uranus by the Doppler shift. Adams and Dunham (1932) discovered carbon dioxide in the Venus atmosphere using the high resolution spectrograph of the 100-inch telescope. For planetary problems high resolution is of prime importance and large telescopes are generally not needed. Because a combination of a relatively small telescope with a powerful spectrograph is useless for stellar work, relatively little high resolution work has been done on planets. The huge telescopes having high priorities for stellar work are used mostly for their large spectrographs. Kiess, Corliss, and Kiess (1960) have alleviated this situation somewhat by use of a small heliostat instrument to feed a big spectrograph. I once used the 16-inch coronograph at Sacramenta Peak to obtain spectra of Venus at 8000 A with a dispersion of 0.6 A/mm. Exposures of 6 hours were required but useful spectra were obtained.

The extension of all of these types of measurements into the infrared spectrum has been of great value. There are fundamental reasons for the importance of infrared studies which it might be well to look at. The reasons essentially may be stated in this way: the infrared region is to planetary studies as the visible and ultraviolet regions are to stellar studies. Planets might be considered as low-energy stars whose source of energy is external rather than internal. Many of the developments of the astrophysics of stars is applicable to the planets in the infrared spectrum.

Planets emit infrared by blackbody emission, perhaps modified by atmospheric and other absorptions. We may see this by looking at the graph of Fig. 1, which shows the solar radiation curve and blackbody curves at temperatures of 100, 150, 200, 250, 300, 500°K. They are all drawn so as to represent the same total emission. Thus if the planetary albedo is 0.50 for solar radiation the crossover point of the solar curve with the planet's appropriate temperature curve is the wavelength at which the planetary light is half reflected sunlight and half intrinsic radiation. Beyond this point the planetary radiation is largely heat radiation. By measuring the radiation in this region one may establish the temperature of the planet. Moreover it is quite possible to investigate temperature distributions over the surface and, if appropriate atmospheric absorptions are present, vertically as well. Realistic measurements of planetary temperatures were first made in the 1920's by Coblentz and Lampland at Lowell and by Pettit and Nicholson at

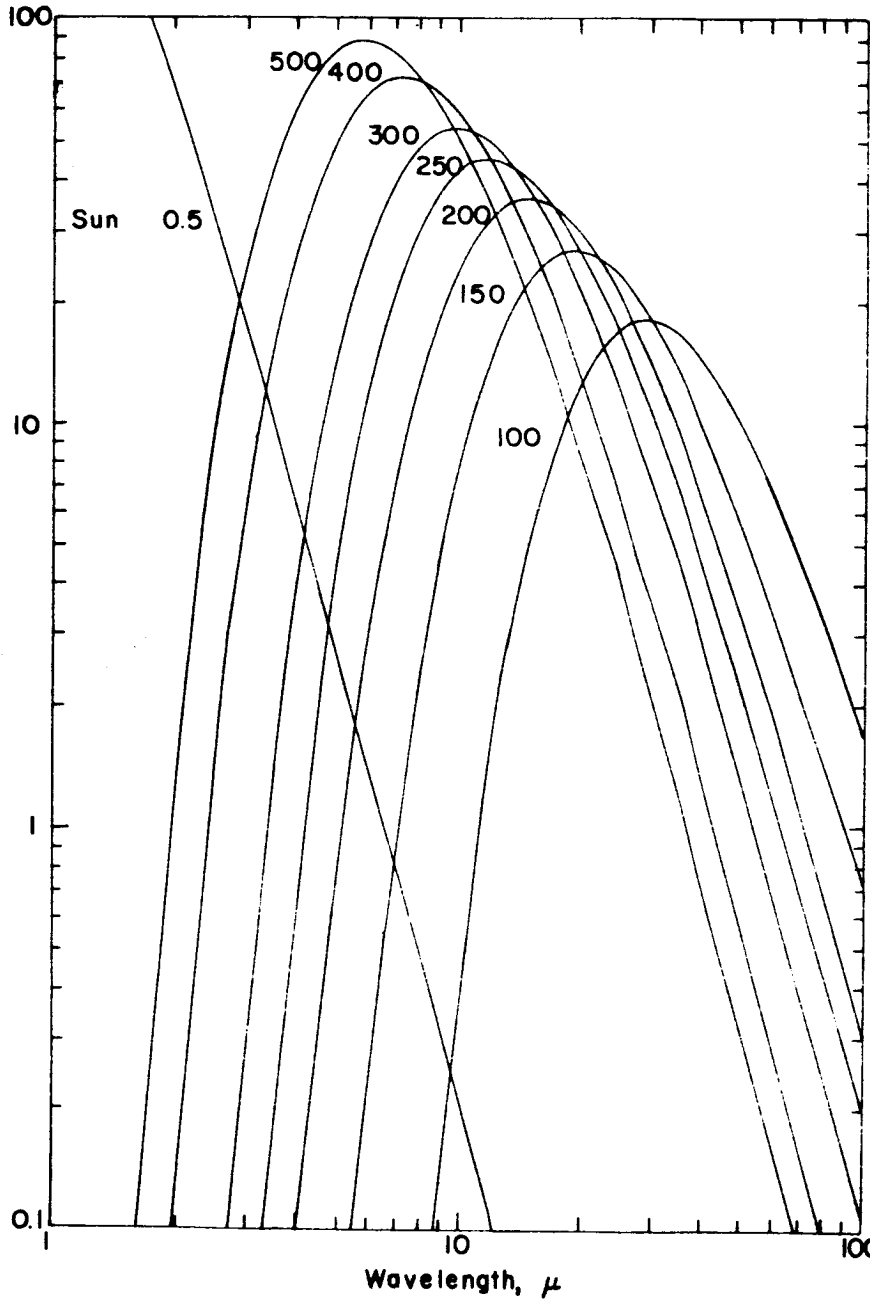

100
10
1
0.1
500
400
300
250
200
150
100
Sun
0.5
1
10
100
Wavelength, μ

Fig. 1

Mt. Wilson. They used infrared-opaque filters to separate the reflected sunlight from the total radiation thus leaving the infrared. Recent work by J. Strong and myself has made use of infrared-transmitting filters and the technique of chopping and has yielded somewhat more sensitivity than with the older methods.

The stable states of planetary matter are solid, liquid, and gas, with molecules being the common constituent rather than atoms. The ionization and dissociation energies of molecules are so much above the planetary thermal energies that this is so. Energies of molecular excitation lie below the ionization and dissociation range, and therefore are found in the thermal range. Stars have higher thermal energies; atoms and ions are found; and their excitation energies are in the visible and ultraviolet.

The molecular excitation states that are of most interest are the vibration and rotation states. Transitions between vibration states are found in the near infrared (1 to 20μ) and transitions between rotation states are found in the far infrared (20μ to 1 mm). These transitions may produce absorptions in the planetary spectrum, and offer the possibility of identifying the molecular constituents of atmospheres and surfaces. Not all molecules, however, absorb in the infrared. Homopolar molecules, e.g., N_2, O_2, H_2, have no electric dipole moment and any absorptions they might have would be due only to the magnetic dipole or electric quadripole moments, which give much lower transition probabilities.

Because atmospheric temperatures are generally lower than surface temperatures the bands appear in absorption. This is not necessarily so, and we may expect to find emission from molecules that are concentrated in the high atmosphere, where temperatures are higher than the surface temperature. Thus is offered the possibility of finding the vertical temperature structure of the atmosphere.

Another way in which the atmospheric temperature may be found is from the molecular population distribution among the various rotational states of molecules. This has been used first by Adel (1937) and subsequently by Kuiper and Chamberlain (1956) to ascertain an atmospheric temperature of Venus. The method is essentially this: in a vibration-rotation band the lines having maximum intensity are selected. They are, of course, assigned roation quantum numbers J. Then $T = (Bhc/2k)(2J_{max} + 1)^2$ where B is the rotational constant, h is Planck's constant, c is the velocity of light, and k is the Boltzmann constant. This technique was refined by Kuiper and Chamberlain to consider all of the lines of the band and avoid the difficult task of selecting *the* line of maximum intensity.

Liquids will have no rotation spectrum but they do have vibration spectra. Crystalline solids have lattice vibrations, which give rise to "residual rays". Very strong lattice vibrations, and also vibration within molecular groups of the crystal, give rise to very low emissivities at specific wavelengths and these in turn show up in the thermal emission. So far no residual rays have been found in planetary spectra except that the emission from White Sands National Monument has shown the residual rays characteristic of gypsum. Silicates have strong bands at 8.8μ. These have been looked for on Mars

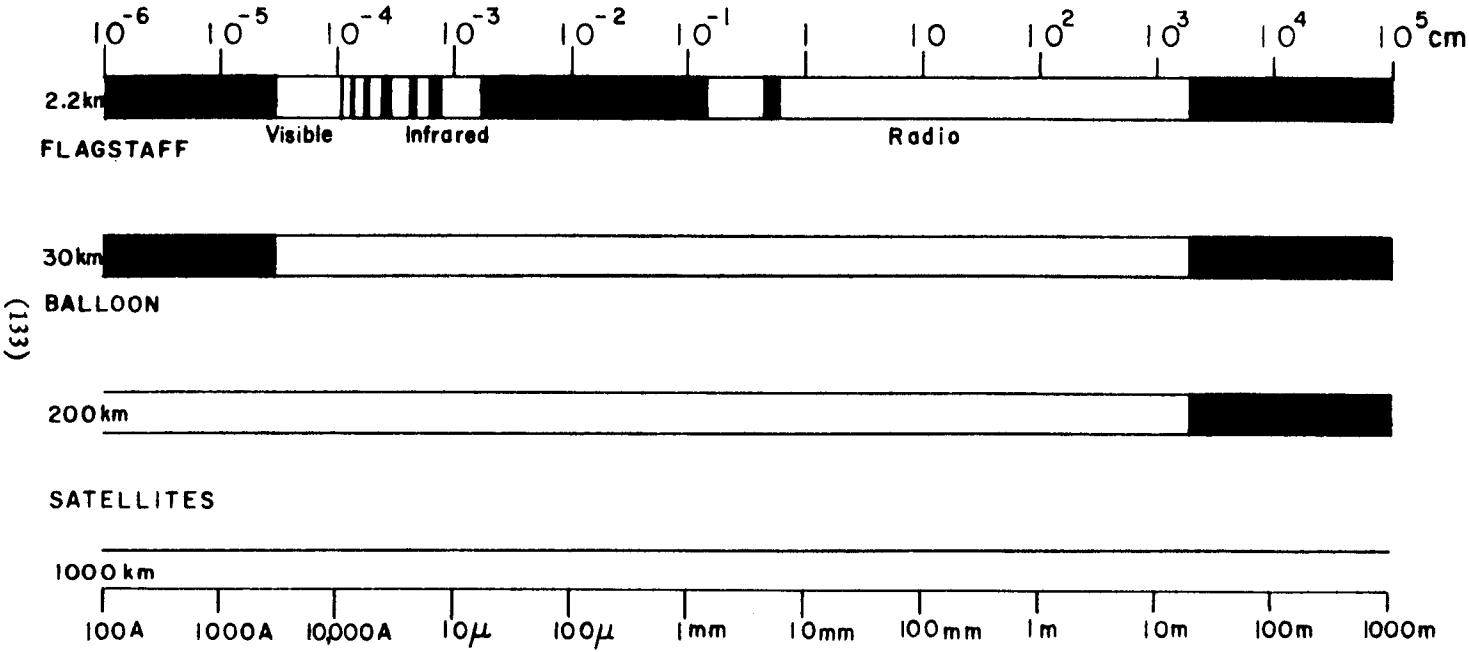

10^{-6} 10^{-5} 10^{-4} 10^{-3} 10^{-2} 10^{-1} 1 10 10^{2} 10^{3} 10^{4} 10^{5}cm
2.2km
FLAGSTAFF
Visible
Infrared
Radio
30km
BALLOON
200km
SATELLITES
1000km
100A 1000A 10,000A 10μ 100μ 1mm 10mm 100mm 1m 10m 100m 1000m

Fig. 2

and the moon but have not been found. Water and ice have strong absorptions at 1.5μ. Kuiper (1952) has found these in the Martian polar cap and Saturn's rings.

The extension of measurements into the infrared is limited by the transmission of the atmosphere, as in this graph (Fig. 2). Most of the absorptions are caused by water vapor. This absorption and that due to other molecules will be greatly reduced by balloon-borne observatories.

The extension of observation into the microwaves is already giving a wealth of information about temperatures and non-thermal emissions. Microwaves are more likely to arise from deeper levels in a planetary atmosphere and they offer the possibility to determine the surface temperature on clouded planets. The microwaves may even arise from somewhat beneath the surface. From a comparison of infrared and microwave temperatures one can learn about the microwave absorption of the surface material.

Planetary Surfaces

The only planets that we see the surface of are Mercury, moon, Mars and probably Pluto. We have substantial knowledge of the surfaces of the first three of these. Visual or photographic observation of these reveal that there is differentiation of material—there being a variation of albedo and color over the surface. Considerable surface ruggedness is apparent on the moon, but the surface of Mars is relatively smooth, and Mercury is too poorly observed to tell.

Photometric measurements give some more insight into the character of the surface. The graph of Fig. 3 gives the polar diagram of the light scattered by these three objects and also Venus. Extremely rough surfaces have a strong backward scattering component owing to hiding of shadows of objects when illuminated along the line of sight. The moon and Mercury have such backward scattering. The significance of the lunar curve has been examined by numerous people, the most recent being van Diggelen (1958). He found that the lunar curve could be explained by assuming hemi-elliptical cups in the surface which was covered by a volcanic ash, but undoubtedly other explanations are tenable. A smoother surface is indicated for Mars from the curve.

Photometric studies also reveal differentiation of the surface by intensity and color variation associated with rotation. The lunar curve above reveals that the moon's light is not symmetrical with phase as a result of the uneven distribution of maria and uplands. Mars exhibits a color variation of 0.15 mag which is related to rotation of the planet. Pluto has been found to have a 0.1 mag variation in its total light. The effects of rotation of many minor planets has similarly been studied.

Polarimetric studies of the planets have also been fruitful. A. Dollfus has shown that polarimetry can tell if a surface is 1) smooth, 2) formed by more or less distinct grains, 3) composed of more or less polished grains, 4) absorbing to a certain degree, or 5) composed of completely absorbing grains.

Figure 4 shows the polarization curves of Mars, Mercury, and the moon. The general shapes of all three are the same. The waxing moon is different

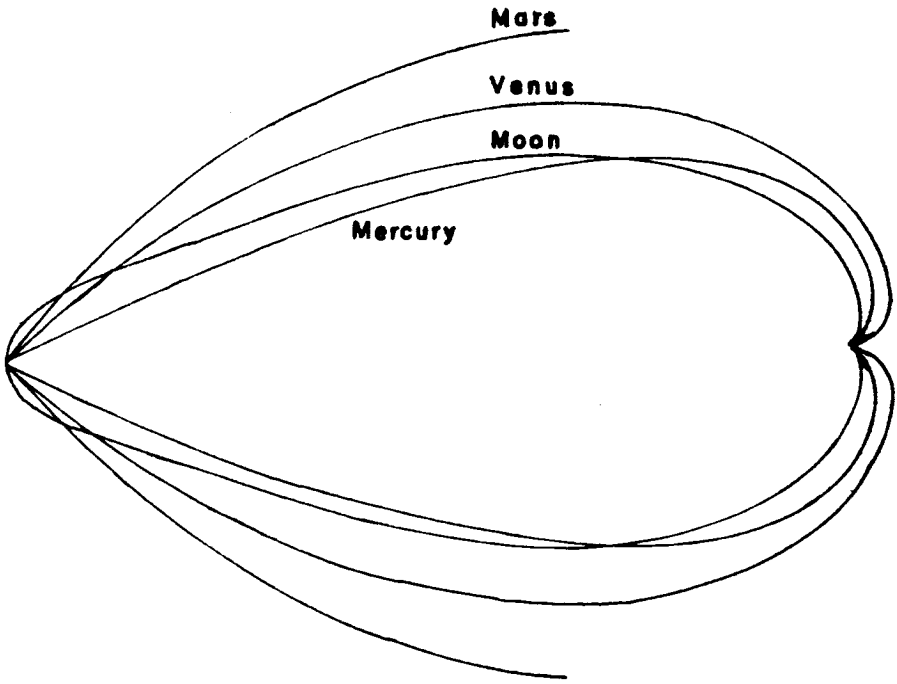


Fig. 3

than the waning moon because of the differences in the two hemispheres of the moon. The curves indicate that all three objects have dark grains on their surfaces. We will come back to the polarimetric measurements later in discussing life on Mars but now we will consider the surface temperature.

E. Pettit and Seth B. Nicholson (1930) at the Mount Wilson Observatory did the most significant work on the temperature of the moon. Many of the observations were made at the 100-inch telescope. For lunar work one thermocouple receiver was placed on the moon; the other was exposed to the sky. The thermocouples were exposed both with and without the cover glass. The difference, after correcting for reflection losses, gave the heat radiation beyond the cut-off of the cover glass at about 5μ. Calibration of the sensitivity of the equipment was made from observations of stars of well-known temperatures. Correction for absorption by the atmosphere was made in two steps. The fraction absorbed in regions where the atmosphere is opaque is calculated for each temperature. The part absorbed in the transparent regions was determined by plotting the log of deflection against the air mass during the observation.

Pettit and Nicholson established the maximum observed temperature of the lunar surface (407°K at the center of the full moon) and the minimum (120°K at the center of the dark hemisphere). However, they found that the temperature of the subsolar point was not constant, but varied according

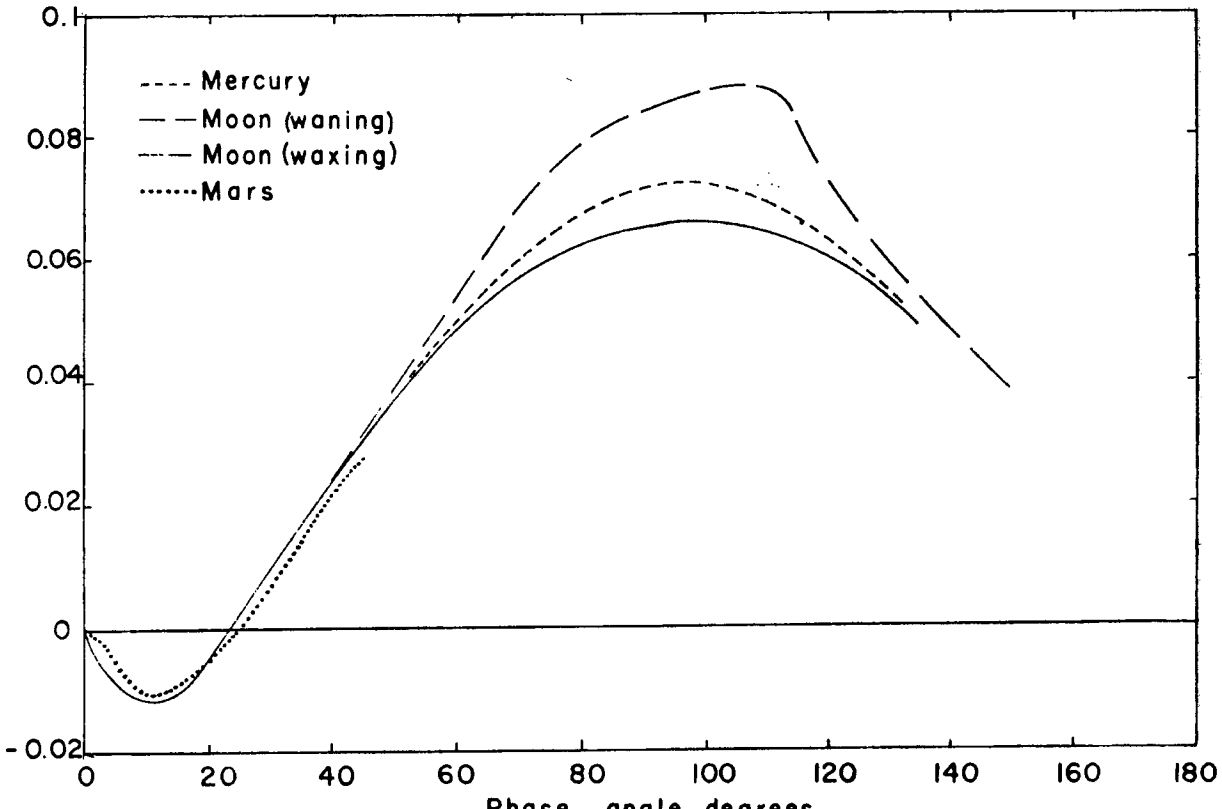

0.1
0.08
0.06
0.04
0.02
0
-0.02
0
20
40
60
80
100
120
140
160
180
Phase angle, degrees
Mercury
Moon (waning)
Moon (waxing)
Mars

Fig. 4

to its position on the disk, as shown in the polar diagram of Fig. 5, which gives the total energy radiated as a function of the angle from the normal to the surface. They found that although the temperature of the subsolar point was 407°K at full moon, it was only 358°K at quarter-phase. This phenomenon is a result of the roughness of the surface and will be discussed more later.

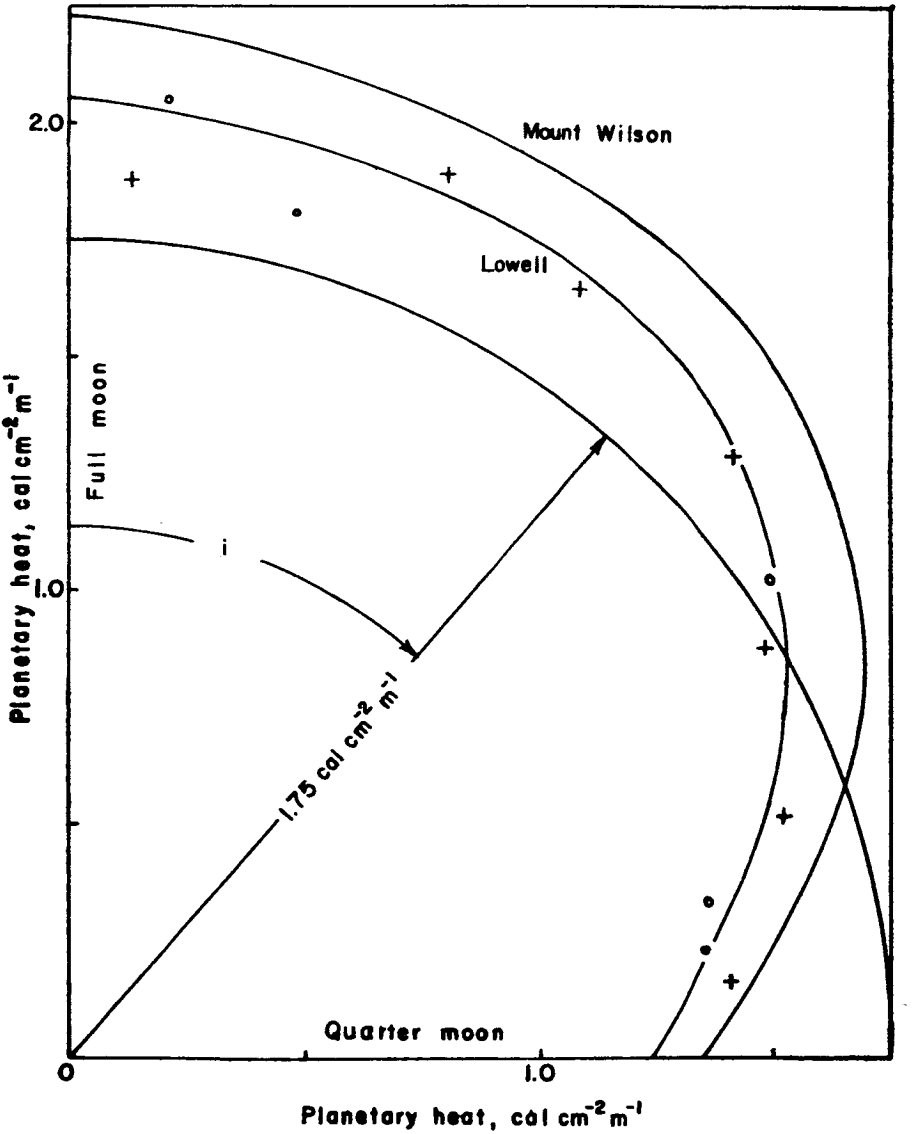


Fig. 5

At Lowell Observatory we have been measuring the moon using the most modern infrared techniques. We use the procedure of chopping rather than opposing junctions to eliminate drift; a narrow-band filter at 8.8μ where the atmosphere has high transparency. The Lowell measurements are also shown in the figure, and compared to Pettit and Nicholson we find about 5% less emission. The temperature of the mean spherically emitting surface is 389°K.

The temperature at "midnight" is much less well determined. We have measured this temperature on several nights at Lowell (Sinton, 1959) and found 122± 5°K. However, we do not have observations at various obliquities in order to make an analysis as we did for the subsolar point.

Perhaps the most important single piece of information comes from measurements of the course of lunar temperatures during a total eclipse. Pettit and Nicholson (1930) measured a point only 48″ from the south limb during the eclipse of June 14, 1927. Pettit (1940) measured an area near the center of the moon's disk during the eclipse of Oct. 27, 1939. John Strong and W. M. Sinton (unpublished) made drift curves across the moon's face during the eclipse of July 26, 1953. All of these measurements are characterized by an extremely rapid change in temperature during the penumbral phase of the eclipse. In the hour that it takes a point on the moon to move through the penumbra of the earth's shadow, the temperature plummets from 374°K (the effective mean-spherical subsolar temperature obtained by Pettit and Nicholson is given) to about 200°K or lower at the beginning of totality. The Mount Wilson observations show that the temperature then proceeds to decline slowly until sunlight again falls on the moon at the end of totality. The temperature rises very rapidly and by the end of the eclipse it is back to its starting point.

The measurements of Strong and Sinton also show an interesting effect when comparing the measurements of the limb to those of the center. They chose five points from their drift curves corresponding to five lunar regions. Two of the regions of Fig. 6 (I and V) were 2′ from the east and west limbs. Region III was in the center of the disk and regions II and IV were between I and V respectively and the center. The central regions, II, III, and IV, followed quite well the curve of Pettit for the 1939 eclipse and achieved a temperature of 213°K at the commencement of totality. The limb regions, on the other hand, were much colder at the beginning of totality (190°K) and compare with the 174°K achieved at a point closer to the limb in the 1927 eclipse. It may be argued that these limb regions were colder to begin with, but a little consideration will show that they have lost a much larger fraction of their total heat than the central regions. The significance of the limb effect will be discussed later but first we need an understanding of thermal conduction into the moon's surface.

We now proceed to outline a theory of the surface temperature of the moon following the work of Epstein (1929), Jaeger (1953a), and Wesselink (1948). We will then compare this model with the observational data already presented.

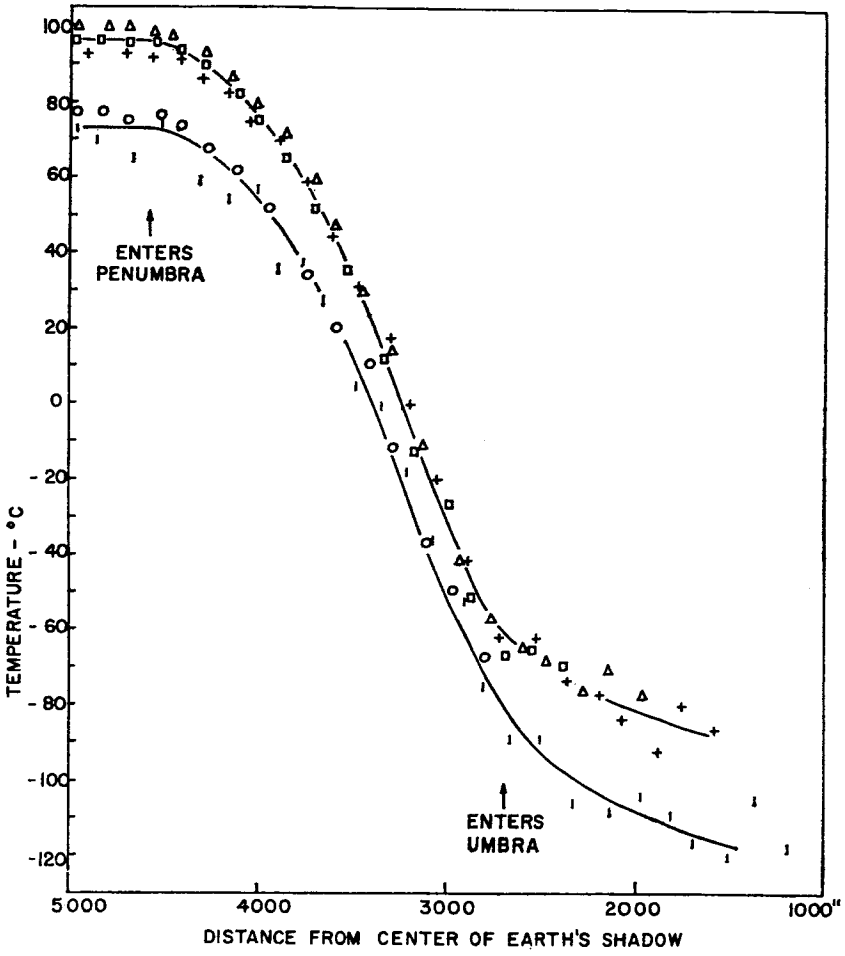


Fig. 6

The surface of the moon is imagined as a homogeneous, semi-infinite solid material characterized by a thermal conductivity k, a density ρ, and a specific heat c. In such a material the temperature T at any depth x below the surface and at time t may be found from the one-dimensional heat-conduction equation,

$$\partial T/\partial t = (k/\rho c)(\partial^2 T/\partial x^2).$$

The heat flux outward at any point is given by the equation,

$$F = k(\partial T/\partial x).$$

In the case of the moon, the surface receives heat I from the sun and

eventually reradiates this to space. The boundary condition that must be obeyed is

$$\sigma T_0^4 = I + F_0,$$

where σ is the Stefan-Boltzmann constant, and the subscript zero signifies the surface. We assume that during the half period when the sun is shining on the surface the insolation is

$$I = (G/r^2)(1 - A)\cos 2\pi(t/P),$$

where G is the solar constant, r is the distance from the sun in astronomical units, P is the period of rotation of the moon, and A is the albedo of the surface. During the other half, when that part of the lunar surface is turned from the sun, I is zero.

The solution of the equations can not be found by Fourier series or other analytic methods because of the non-linearity of the radiation term in the boundary condition. Wesselink has found one solution by a method of numerical integration. It is, however, tedious to do by hand but can probably be readily adapted to machine computation.

Jaeger, on the other hand, has found an easier iterative method wherein a solution is first assumed and then modified according to where it is seen to fail.

For a given environmental situation involving a definite insolation and rotation period, there are a family of solutions which are given parametrically by the thermal inertia $(k\rho c)^{1/2}$. The thermal inertia measures the resistance of the surface to a change in temperature; the greater the value of $(k\rho c)^{1/2}$, the smaller is the temperature variation.

In Fig. 7 we present the curves of surface temperatures that have been computed. The thermal inertias are given in calories, centigrade degrees, and cgs units. The curves with inertias of 0.001, 0.002, 0.004, 0.008, and 0.05 have been taken from Jaeger, and the two remaining curves have been computed by the author, one by each method. In computing the curves, the solar constant was assumed so as to make the maximum temperature 370°K. In the previous section we saw that this should be raised to 389°K. This can readily be done by means of the homologous transformation that has been given by Wesselink.

If the following substitutions are made,

$$T = \alpha\, T_* \qquad x = \beta\, \alpha^{-3}\, \chi_*$$
$$(k\rho c)^{1/2} = \beta (k\rho c)^{1/2}_* \qquad I = \alpha^4 I_*$$
$$t = \beta^2 \alpha^{-6} t_* \qquad F = \alpha^4 F_*,$$

then the starred quantities are also solutions to the heat conduction and boundary equations. Since these transformations can change the time scale as well as the insolation, the solutions can be used for other atmosphereless planets having different rotation periods and distances from the sun. For our present case, to correct the temperature scale, we do not desire a change in time scale, and therefore, $\beta^2 = \alpha^6$. The ratio of the temperatures, 389/370, gives $\alpha = 1.05$ and $\beta = 1.16$. The curve given in Fig. 8 is the

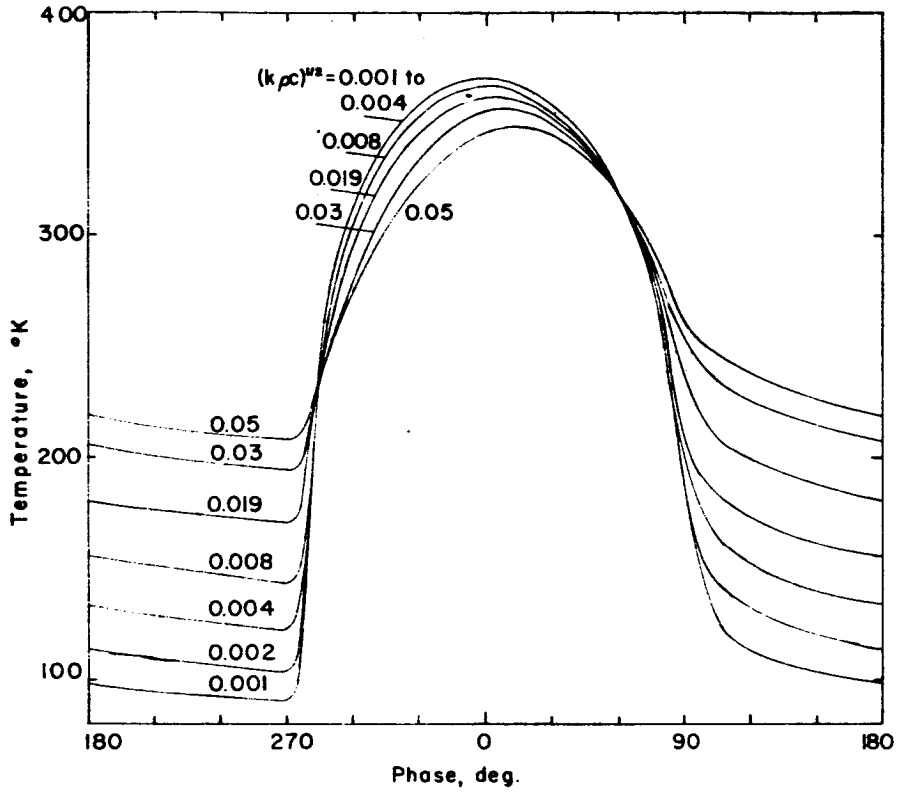


Fig. 7

transformed curve from the previous graph where the thermal inertia was 0.002. It may be seen that this is a reasonably good fit to the experimental points. However, the critical "midnight" point is not accurately determined and smaller inertias are indicated by eclipse observations.

The agreement shown in the figure demonstrates that our model of the moon's surface works tolerably well and that the thermal inertia of the surface is very low. A search through tables of constants reveals that ordinary rocks such as granite and basalt have inertias near 0.05. Dry soils and sands range between 0.01 and 0.02. Pumice has one of the smallest inertias of natural materials but it is still of the order of 0.004. Even pumice does not offer an adequate explanation of the low inertia of the lunar surface.

But the reason that granulated natural materials do not have a sufficiently low conductivity is that their conductivity is greatly raised by air in the interstices. Smoluchowski (1910, 1911) investigated the thermal conductivity of powders at gas pressure down to a good vacuum. At a high vacuum the conductivity for an average of 11 grain sizes smaller than 0.1 mm was $k = 3 \times 10^{-6}$cal $cm^{-1}sec^{-1}(°C)^{-1}$. If we combine this with a density of 1.5 and a specific heat of 0.2, typical of most rocks and minerals, we

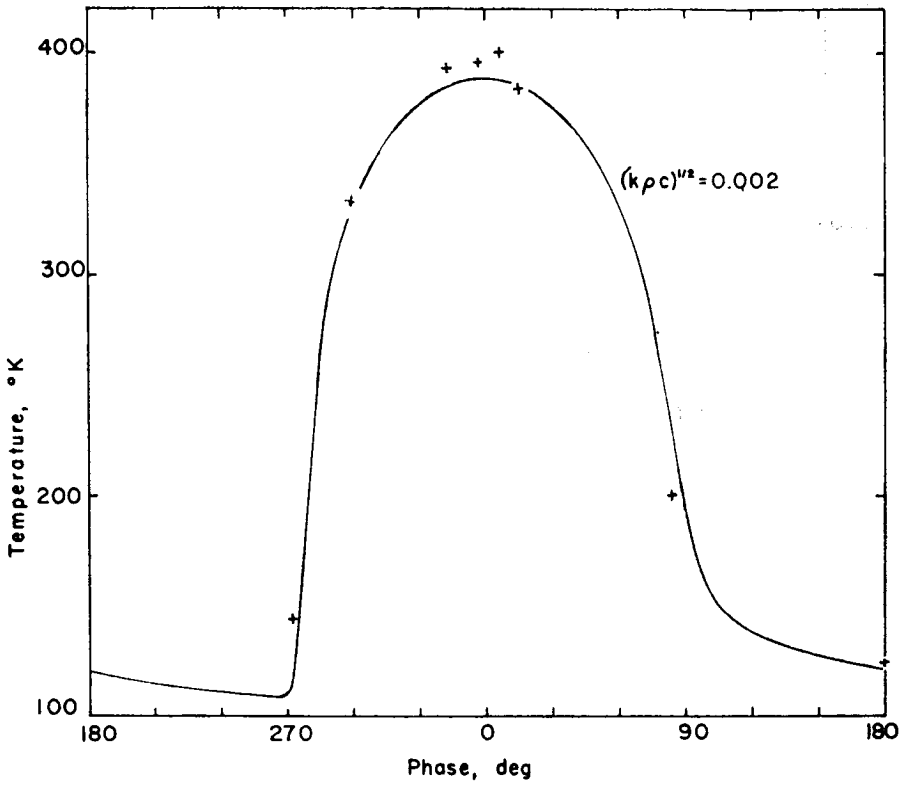


Fig. 8

find $(k\rho c)^{1/2} = 0.001$. Powdered material or dusts in a vacuum therefore appear to offer a satisfactory explanation of the very low thermal inertia of the lunar surface.

Beneath the surface a thermal wave is propagated downward and satisfies the conduction equation. Here, we may use Fourier analysis and the solution found at the surface is resolved into such a series. The surface variation is expressed as

$$T(O,t) = \sum_{n=o}^{\infty} T_n \cos(2\pi nt/P + \epsilon_n).$$

At any depth x below the surface, the temperature is

$$T(\xi,t) = \sum_{n=o}^{\infty} T_n e^{-2\pi\xi n^{1/2}} \cos[2\pi nt/P) - 2\pi n^{1/2}\xi + \epsilon_n].$$

where $\xi = x/l$, and l is the wavelength of the fundamental wave given by $l = 2(\pi Pk/\rho c)^{1/2}$. The above equation does not contain the diffusivity $k/\rho c$ explicitly but only through l so if we concentrate on ξ, we do not need to know the diffusivity. Each component wave is propagated into the

material with a strong attenuation cooefficient given by $2\pi n^{1/2}$. Since the wavelength of each component is $n^{-1/2}$, in one wavelength the wave is reduced to $e^{-2\pi}$ of its amplitude.

Using $(k\rho c)^{1/2} = 0.001$, we evaluate the T_n and ϵ_n and find the values given in the table. We note that the fundamental wave lags the phase of the moon by 6°. This phase shift is produced by the slow decay of the lunar temperature during the "night".

Fourier Series Parameters for Lunar Surface

n	T_n	ϵ_n
0	210°K	—
1	157	−6°
2	34	+6
3	30	159

By use of the homologous transformation any solution that has been found for the moon becomes a solution of Mars but for different values of the thermal constants. For any two solutions for Mars and the moon that are related through the above equations, the temperature of Mars is 0.855 that of the moon, i.e., $\alpha = (r_\bullet/r)^{1/2}[(1 - A)/(1 - A^*)]^{1/4}$, $\alpha = 0.855$, and the $(k\rho c)^{1/2}$ of Mars is 0.12 of that corresponding to a solution having the same phase lag on the moon, i.e., $\beta = \gamma^{1/2}(r_\bullet/r)^{3/2}[(1 - A)/(1 - A^*)]^{3/4} = 0.12$. Because the distance of Mars from the sun varies considerably, these factors are not constant. The above values are derived for the date of July 20, 1954 when Strong and Sinton (1960b) made their measurements.

Three theoretical curves are given in Fig. 9 along with the observed temperatures. The curve with $(k\rho c)^{1/2} = 0.001$ was obtained by trans-

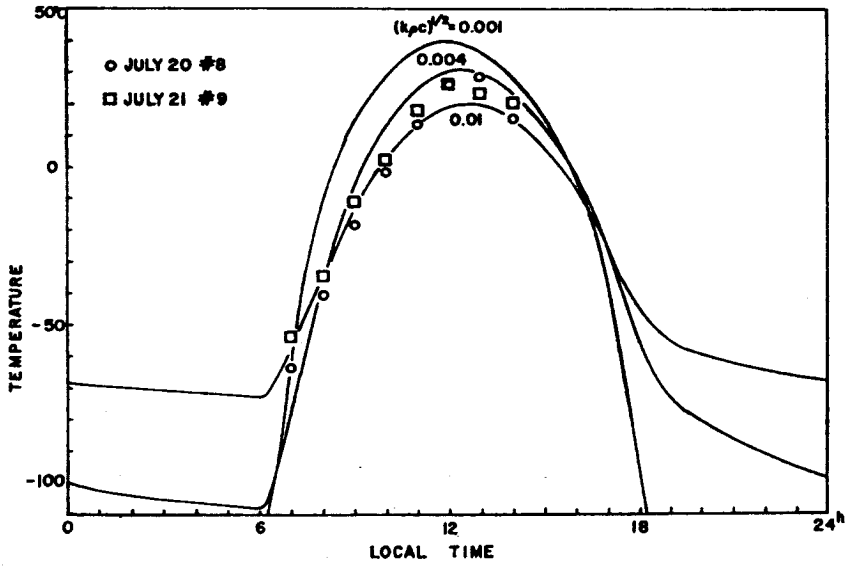


Fig. 9

formation from Jaeger's curve for the moon with $(k\rho c)^{\frac{1}{2}} = 0.008$, while the curve for 0.004 was obtained from the curve calculated for the moon, using Wesselink's method, and the curve with $(k\rho c)^{\frac{1}{2}} = 0.01$ was calculated by Jaeger's method for Mars.

The application of this theory, which does not include the action of the atmosphere, to a planet with an atmosphere even as tenuous as that of Mars is somewhat questionable. The differences between the predicted curves and the observed values can yield useful information about the effect of the atmosphere.

The atmosphere enters into the problem primarily through the boundary conditions expressed in the boundary equation. Two additional terms are necessary: the heat conducted to or from the surface by the atmosphere and the heat returned to the surface by radiation from the atmosphere. It is assumed that, for the case of Mars, atmospheric attenuation of the incident solar radiation is negligible. The boundary equation becomes

$$H + \sigma T_0^4 = I + F_0 + R,$$

where H is the heat lost by conduction to the atmosphere and R is the heat gained by radiation back from the atmosphere. The main effect of the additional terms is to reduce the amplitude of the temperature-curve. During the night, both H and R supply heat to the surface, while in the day H takes heat away from the surface while R continues to supply heat. If the atmospheric effects are small, these terms will probably not affect the phase appreciably. In the case of the earth these considerations seem to apply. The amplitude of the surface temperature-curve is much reduced, while the phase of the curve seems to be fairly well given by the thermal constants of the surface.

From the previous figure it appears that these effects also play a role in the surface temperature of Mars. The observational data appear to agree in phase with the curve for $(k\rho c)^{\frac{1}{2}} = 0.004$, while the amplitude is more nearly that of the curve with 0.01 for these constants. It may thus be possible from a further analysis of the observed curve to determine something further about the meteorology of the planet. It is emphasized that the only parameter which has been adjusted in comparing the theoretical with the observed values is $(k\rho c)^{\frac{1}{2}}$. The maximum temperature at a favorable opposition for a desert area near the equator of the planet appears to be close to 25°C, and for a dark area it is about 8° hotter. It should be emphasized that these temperatures are surface temperatures and are certainly higher than the air temperature near the ground. The minimum temperature is more uncertain, but is probably in the neighborhood of −70°C for a bright area on the equator at perihelion.

Terrestrial rocks have inertias near 0.05 while sandy or dusty solids range between 0.01 and 0.02. The presence of moisture appears to raise the inertia. A dark volcanic ash at the Grand Falls of the Little Colorado River, near Flagstaff, Arizona has $(k\rho c)^{\frac{1}{2}} = 0.01$ and a measured surface temperature change from 3° to 72°C through a day near the summer solstice. The ob-

served lag of the maximum for this ash was a little over an hour and this is the lag expected of a surface with an inertia of 0.01. On the other hand the air one meter above the ground did not attain its maximum temperature of 30.5°C until three hours after the local noon. We may conclude from these observations that the lag is not sensibly affected and the amplitude is greatly moderated by the presence of an atmosphere. For the generally dry, dusty surface of Mars an inertia of 0.004, a lag of about one-half hour, and a total variation near 100°C appear quite reasonable.

The analysis of the diural temperatures may be applied to the minor planets, satellites, and Pluto by use of the homologous transformations. All that need be known or assumed are the albedo, rotation period, distance from the sun and the thermal inertia.

The heat conduction equations are applicable to the temperature change during a lunar eclipse. Here, of course, the changes are much more rapid and I decreases from its largest value to zero within about an hour. Depending on the magnitude of the eclipse, the duration of totality may last up to 3 hours, and then emergence occurs in the same time as the initial penumbral phase.

Solutions may be found by numerical integration and many have been found by Wesselink and Jaeger for the homogeneous one-constituent model. A curve has been calculated by Wesselink for an inertia of 0.001, and reasonable agreement is found with the eclipse observations of Pettit in 1939. Pettit has assumed that the temperature before the eclipse was 370°K and the equipment was not otherwise calibrated. If we want to take 389°K for this temperature, we may use the transformation given before. The temperatures will be raised by 1.05 and the inertia will be raised to 0.0012.

In Pettit's data the general range of the variation is correct, but during totality the observed curve is much flatter than the computed. Jaeger and Harper (1950) have resolved this disagreement with a two layer model. They have found for a thickness of 0.17 cm of dust with $(k'\rho'c')^{1/2} = 0.00097$ overlying a base material having $(k\rho c)^{1/2} = 0.01$ that better agreement with Pettit's observed curve is obtained. The data of Strong and Sinton appear to show greater slope in the umbral portion of the eclipse, but they employed a rather large aperture (3'.8) and smoothing effects may possibly be responsible for this.

Shorthill, Borough, and Conley discovered during the total eclipse of the moon on March 13, 1960 that the floor of the crator Tycho did not cool as much as the surrounding area, but remained some 40 to 60° warmer. They found that other rayed craters, specifically Copernicus and Aristarchus, also remained warmer than their surroundings. The author made scans across Tycho during the eclipse of Sept. 5, 1960. Fig. 10 shows two of the scans; one is before the traversed chord of the moon is completely shadowed. The time interval between crossing the western limb and crossing Tycho was calculated and compared to the tracings obtained before totality. Such a comparison is shown in the figure. The agreement is good, showing that the warmer region is indeed Tycho.

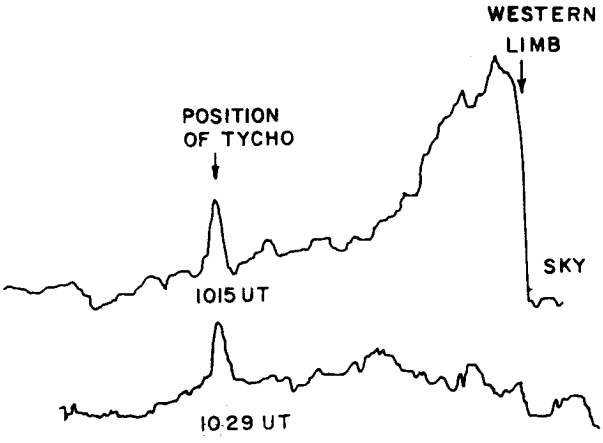


Fig. 10

The next graph (Fig. 11) shows the temperatures observed for Tycho as dots; crosses designate the temperatures of the region around Tycho found by taking the mean of the deflections on either side of the bump. Several theoretical curves, which assume different combinations of dust and rock, are taken from papers of Jaeger (1953) and Jaeger and Harper (1950), and they are also shown in the figure. The curve for thick dust agrees well with measurements of the Tycho environs. The measurements of Tycho, itself, agree well in general shape with the curve for 0.5 mm of dust on top of rock; they show little change during totality. A similar curve, if computed for a slightly thinner dust layer, would fit the points better. The computation of this curve requires numerical integration.

However, the thickness of the dust layer that would fit the observations better can be determined by assuming that, after totality has commenced, the heat radiated from the surface is equal to the heat conducted through the layer. This is expressed by

$$\sigma T^4 = (T_0 - T)k/d,$$

where d is the thickness of the layer. Jaeger and Harper assumed $k = 2.8 \times 10^{-6}$ in cgs units. The temperatures during totality, which they computed by numerical integration, are adequately derived from the above formula if T_0 is 0.73 of the pre-eclipse temperature. From this formula and the observed eclipse temperature of Tycho, the thickness of the dust layer is 0.3 mm.

Another possible model was considered, that of bare boulders lying in a field of dust, and the curve for this model is also shown. The agreement with the dust-layer model is decidedly better. In the case of the two points that are of lower temperature, the scans very likely crossed only the edge of Tycho.

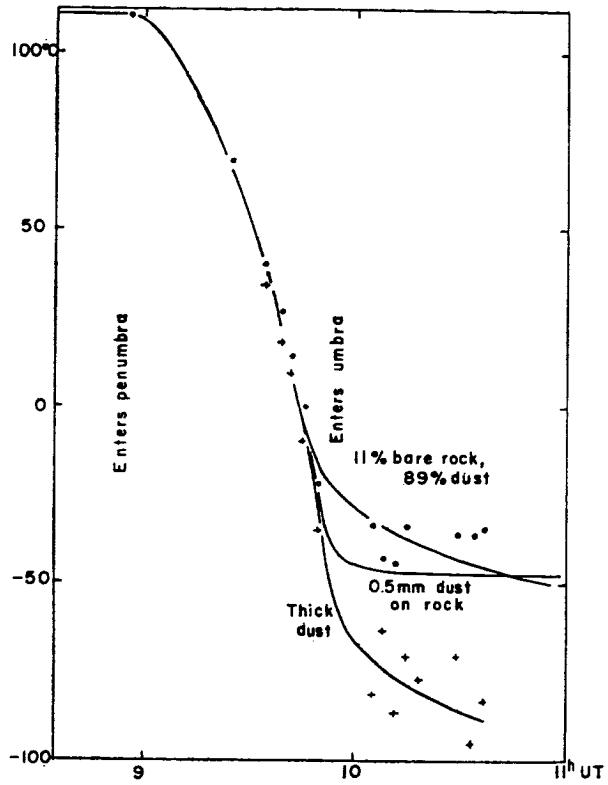


Fig. 11

During the eclipse of March 2, 1961, 8 other rayed craters were measured in addition to Tycho, Copernicus, and Kepler. All of these except one were warmer than their immediate surroundings.

The first microwave observation of thermal radiation from the moon was made by Dicke and Beringer (1946). They made only one measurement at 1.23 cm wavelength. This was followed by extensive measurements of Piddington and Minnett (1949) at the same wavelength. Their antenna beam width at half-power points was $\pm$ (3/8)° and so it averaged over most of the lunar disk. Calibration of the antenna-receiving system was made by using a resistive termination of a wave guide which could be heated. They found that the equivalent antenna temperature followed a nearly sinusoidal variation which was given by

$$T = 39.0 + 6.6 \cos [(2\pi/P)t - \pi/4].$$

They have reduced their measurements to an average along the equator of the moon. Using the antenna beam width and an assumed lunar emissivity of 0.90 they have corrected their temperatures to the variation at the equator.

They obtain for the equatorial temperature,

$$T = 249 + 52.0 \cos [(2\pi/P)t - \pi/4].$$

A number of later investigators have followed their lead and the lunar temperature has been observed in a wide range of wavelengths. Where a monthly variation has been found it has nearly always been sinusoidal. Constants have been found in the following relation for the equatorial temperature.

$$T = T_0' + T_1' \cos [(2\pi/P)t - \Psi].$$

Gibson had a beam width of only 0.2° and found definite asymmetry in drift curves across the moon. He also found that the periodic variation of lunar temperature was definitely non-sinusoidal and was similar to the theoretical curves of Jaeger that are given in the next section.

W. Sinton (1955) observed with essentially optical techniques in a band centered at $\lambda = 0.15$ cm. His beam width included the whole moon. Effective temperatures ranged from 120°K near new moon to a mean temperature of 336°K near full moon. Much scatter was present and calibration, which was by means of solar observations, was difficult. But the measurements serve to show approximate agreement with the infrared measurements at 10μ.

Sinton (1956) also observed the total lunar eclipse of January 18, 1954. He found that the initially observed temperature of 300°K dropped to about 170°K during the eclipse. Moreover, the observed temperature variation appeared to lag behind the optical eclipse by an hour. Thus the characteristics of 1.5 mm lunar radiation are somewhere between the infrared and the microwave behavior. For a long-period variation in insolation such as throughout a lunation, the apparent temperature variation is large and nearly in phase while for a rapid variation as during an eclipse there is a small lag and the range of variation is reduced from that which is observed at 10μ. On the other hand microwave observations of eclipses have always found no change in apparent temperature (Gibson, 1958, 1961; Mezger and Strassl, 1959).

At Lowell Observatory the author, during the eclipse of March 13, 1960, attempted to confirm the earlier 0.15 cm eclipse observations. The observed area was now restricted to 6′, and measurements were made of Mars Tranquilitatis. The signal-to-noise ratio, however, was much poorer than in 1954 and many clouds and a more humid atmosphere hindered the observing. These data seem to show no variation during the eclipse casting some doubt on the previously found variation. Until further observations have been made, the earlier data will be used.

The small amplitude of variation and the phase lag of the microwave measures demonstrate that the radiation is arising from below the surface of the moon. We have seen that the surface variation produces a thermal wave that propagates into the moon with a rapid attenuation of amplitude. If we confine our attention to the radiation arising from a thin layer below the surface, the emission will have a phase lag characterized by the time for

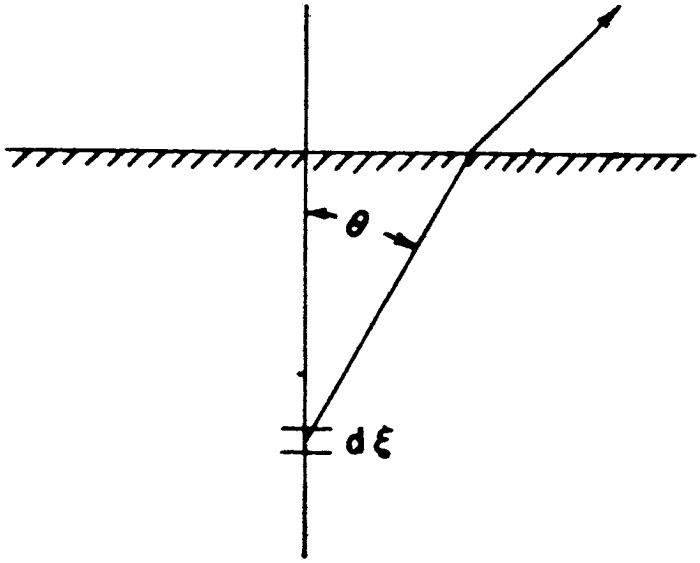


Fig. 12

the thermal wave to propagate to this level and an amplitude variation corresponding to the temperature variation at this level. Actually, the emission comes partly from all levels and the apparent variation is the summation of the emission from these levels.

In Fig. 12 the radiation per unit solid angle and cm^2 arising at angle θ from a layer of $d\xi = dx/l$ in the medium with absorption coefficient $\alpha = \alpha'/1$ and index of refraction n is by Kirchoff's law $n^2B\ (\lambda,T)\alpha'\sec\theta\ d\xi$, where $B(\lambda,T)$ is from Planck's formula. In the microwaves $B(\lambda,T)$ is approximated very well by the expression $CT\lambda^{-4}$ where C is constant. The emission suffers absorption of the amount $1 - e^{-\alpha'\ \xi\sec\theta}$ in reaching the surface. Just above the surface the radiation will be given by

$$I = (1 - R)\alpha'\sec\theta \int_0^\infty B(\lambda,T)e^{-\alpha'\xi\sec\theta}\,d\xi,$$

where R is the Fresnel reflection loss. Using the approximate expression for $B(\lambda,T)$ we find that the apparent microwave temperature is

$$T_a = (I/c\lambda^{-4}) = (1 - R)\alpha'\sec\theta \int_0^\infty T(\xi)d^{-\alpha'\ \xi\sec\theta}\,d\xi.$$

In this discussion, by employing the reduced depth ξ, we have kept our attention upon quantities that are measurable from the emitted radiation. We continue this idea by introducing the mass absorption coefficient $\mu = \alpha/\rho$ from which $\alpha' = (2\mu/c)(\pi P)^{1/2}(k\rho c)^{1/2}$. With this substitution into the above we have only the inertia, which we already know; the period of the moon; the mass absorption coefficient, which does not depend on the state of division of lunar matter; and the specific heat, which may be guessed fairly closely. From tables of specific heats it is seen that nearly all minerals and rocks have specific heats close to 0.2 with a dispersion less than 10%.

The integral expression may be used for the temperature below the surface. Evaluating the integral we obtain for the apparent temperature,

$$T_a=(1-R)T_0+(1-R)\sum_{n=1}^{\infty} T_n[\cos(2\pi nt/P+\epsilon_n-\phi_n)]/(1+2\delta_n+2\delta_n^2)^{1/2},$$

where δ_n has been written for $c/(\mu\sec\theta)(n\pi/P)^{1/2}(k\rho c)^{1/2}$ and $\phi_n = \tan^{-1}[\delta_n/(1+\delta_n)]$. We have already seen from the observations that often only the fundamental term is necessary in the summation. The strong attenuation of the overtone thermal waves makes these unimportant except at very short wavelengths. Furthermore, the maximum ϕ_n that is possible is 45°. Since the fundamental wave already lags 6° from the phase of insolation, the maximum possible lag from full moon for observations of the center of the lunar disk is 51°. We may use either the observations of the ϕ_1, or the observations of the amplitude of the fundamental waves to derive the mass absorption coefficient for different wavelengths. We take $(k\rho c)^{1/2} = 0.001$, $c = 0.2$, and $P = 2.55 \times 10^6$sec. For observations of the center of the disk $\sec\theta$ is 1. If observations extend over the entire visible disk and the dielectric constant is 5, then $(\sec\theta)_{av}$ is only 1.05. In any event, the mean is not larger than 2. To eliminate possible calibration errors and the necessity of assuming an emissivity, we take the observed T_1'/T_0' and set it equal to $(1+2\delta+2\delta^2)^{-1/2} \times T_1/T_0 = 0.748(1+2\delta+2\delta^2)^{-1/2}$. Fig. 13 and Table 3 give the μ's derived by the amplitude and phase lag methods. If the thermal inertia is 0.002, as indicated by lunation measures, instead of 0.001, then the values in the figure are all reduced to about one-half those given.

It is possible that Akabane's observed variation is real if the moon's surface has some resonance near 10 cm, but this is not likely. Microwave observations may possess a systematic error having the lunar period. The new moon must be observed in the daytime and the full moon must be observed at night. Therefore the equipment is subjected to a temperature variation that has the period of the moon.

Sinton (1956) has derived the mass absorption coefficient from his eclipse measurements at 0.15 cm by using some of the theoretical calculations of Wesselink. He has found $\mu = 2.9$ cm^{2}gm^{-1} at this wavelength and this value is also given in the figure.

TABLE III

Mass Absorption Coefficients

Observers	λ	μ, Amplitude Method	μ, Phase Method
Salomonovich	0.8 cm	0.082 cm^{2}g^{-1}	0.107 cm^{2}g^{-1}
Gibson	0.86	0.083	0.119
Piddington and Minnett	1.25	0.107	0.045
Troitsky and Zelenskaya	3.2	<0.03	....
Akabane	10	0.052	0.045
Mezger and Strassel	20.5	<0.008	....
Seeger, Westerhout, and Conway	33	<0.044	....
Davies and Jennison	22	<0.0015	....

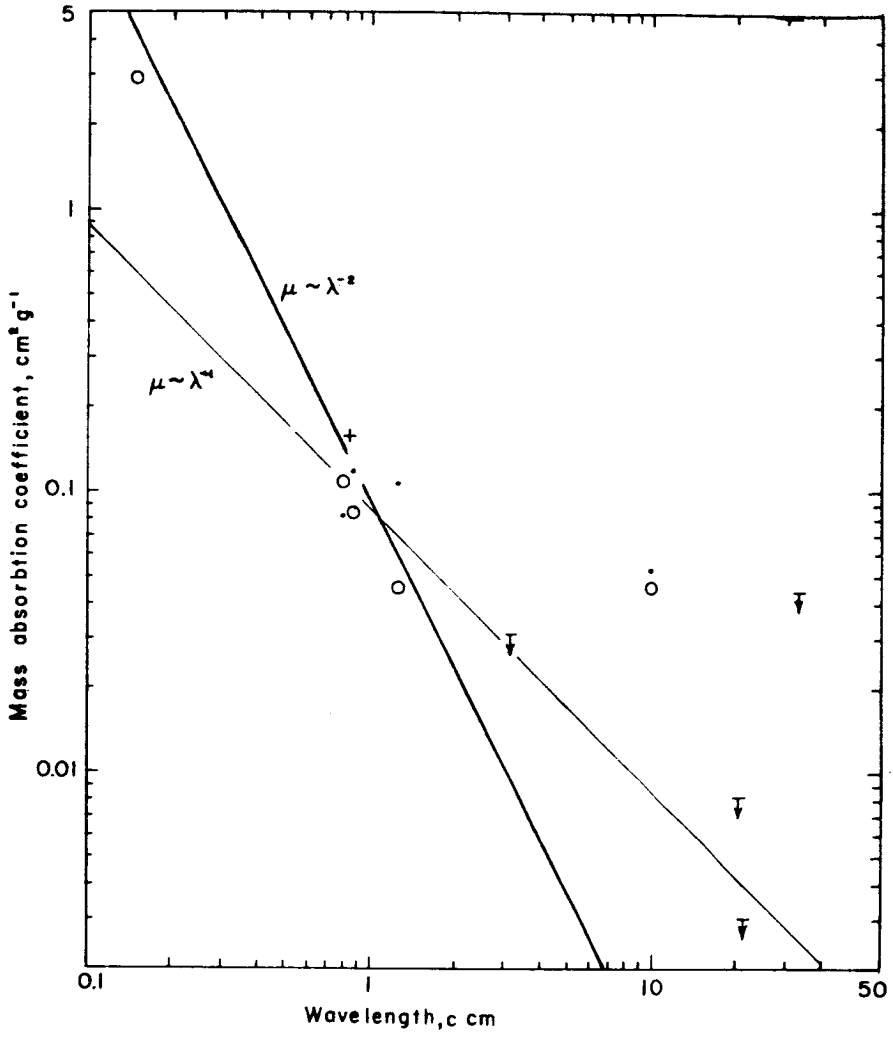


Fig. 13

The absorption in the millimeter region may be produced by two causes. One is the strong crystal lattice vibrations occurring in the far-infrared region. The equation that relates the complex index of refraction to the crystal eigenfrequencies and damping constants is

$$(n - iK)^2 = 1 + \Sigma_i \, (A_i\lambda^2)/[(\lambda/\lambda_i)^2 - 1 + iD_i\lambda].$$

At wavelengths much beyond resonance this expression gives the extinction coefficient k proportional to λ^{-1}. Since k is defined as $4\pi\lambda\alpha$, α and μ are proportional to λ^{-2}.

Another possible cause of the absorption is the electrical conduction of the medium. The absorption coefficient is given by $\alpha = 4\pi\sigma/c\epsilon^{1/2}$ where

σ is the electrical conductivity, c is the velocity of light and ϵ is the dielectric constant of the medium. The conductivity and hense α is found to be proportional to λ^{-1}. Lines with these slopes are plotted in the above figure and the observations are seen to favor the λ^{-2} dependence.

Surface or subsurface temperatures of Mercury, Venus, Moon, and Mars have been measured by microwave telescopes. We give here a table of equatorial surface temperatures of the inner planets

	High	Low	
Mercury	610°K	27°K	(calc.)
Venus	580	500	(uncertain)
Moon	389	120	
Mars	300	200	

Concerning the composition of surfaces of inner planets, one at the moment can give only very general ideas. The cosmically abundant and chemically stable non volatile compounds will generally prevail and one might use as a guide the composition of the earth's crust which is given below.

O	46.6%	Mg	2.1%
Si	27.7%	Ca	3.6%
Al	8.1%	Na	2.8%
Fe	5.0%	K	2.6%
			98.5%

Beside H and He, C and N are much depleted compared to cosmic abundances because they do not form stable, non volatile compounds as does O which is only slightly depleted. Fe is reduced because the bulk of it is in the earth's core.

Spectroscopic tests for various compounds on the planets have largely been negative. Ice has been found in the Martian polar caps and on the rings of Saturn. Kuiper found that the intensity of light reflected from these objects was much reduced beyond 1.5μ in accordance with the reflectivity of ice.

Silicates have been looked for on the moon and Mars through their residual ray bands at 8.8μ. However 20% of silicates might be present and not be found, especially if they are finely divided.

The question of life on any planet is one of the greatest philosophical importance. Mars has long been recognized as the planet other than ours which most likely harbours life. The seasonal variation of the intensity of the dark areas, in phase with the spread of moisture from the evaporating polar cap points to this. A Dollfus has found a seasonal variation in the polarization properties of the dark regions. There is a decrease in polarization at phase angle 25° during the Martian spring in both the northern and southern hemispheres. This change in polarization means that there is a change in the *microscopic properties* of the surface material in the spring.

I have attacked the life problem by looking spectroscopically for the C-H resonance at about 3.5μ. Instead of the two absorptions that I expected, I

found three (at 3.45, 3.58, and 3.69μ). The three are characteristic of the aldehydes, i.e., CHO structures. These bands were found primarily in the dark areas and not in the bright regions. We thus have three good indications of either vegetative or bacterial life. But clearly the life question will not be settled until one makes direct tests on the soil.

Atmospheres I

We begin our study of planetary atmospheres by outlining the theoretical picture of the pressure, density, and temperature structure of an atmosphere. First we may consider, because of its usefulness, the pressure and density dependence on altitude from the base of an isothermal atmosphere. The kinetic theory of gases gives for this that

$$p = p_0 e^{-(mgh/kT)} \quad \text{and} \quad \rho = \rho_0 e^{-(mgh/kT)}$$

where m is the mass of the molecule, g is the force of gravity, h is the height above the base at which the pressure is p_0 and the density is ρ_0, k is the Boltzmann constant and T is the absolute temperature. If the atmospheric gas is a mixture and the gases remain mixed with altitude then $m = \bar{m}$ where $\bar{m}$ is the average mass of the molecule. The quantity $H = kT/mg$ is called the scale height of the atmosphere. For the earth (which, of course is not isothermal) $\bar{m} = 29$ and $T = 290°K$; H is 8.5 km.

The usually encountered lower atmosphere, like the earth, is not an isothermal one, but one which is in adiabatic or near adiabatic convection. For a dry adiabatic atmosphere then

$$p = p_0[1 - \Gamma(h/T_0)]^{[\gamma/(\gamma-1)]} \text{ and } \rho = \rho_0[1 - \Gamma(h^1T_0)]^{1/(\gamma-1)}$$

where Γ has been written for $(\gamma - 1)/\gamma$ $(\bar{m}g/k)$ and is the vertical lapse rate of temperature, i.e., $\Delta T = -\Gamma\Delta h$; and γ is the ratio of specific heats, C_p/C_v of the gas. For diatomic gases γ is 1.4; for polyatomic gases γ is approximately 1.33. The pressures P_1 and P_2 at two levels in the atmosphere are related by $P_1/P_2 = (T_1/T_2)^{\gamma/(\gamma-1)}$.

If the gas contains condensibles, i.e., is not dry, then part of the heat is taken from latent heat of the condensible and a smaller lapse rate applies. For the earth the dry lapse rate is 9.8°/km, while because of the moisture content the actual rate is close to 6 or 7°/km.

This lapse rate changes almost discontinuously to zero at the tropopause at the base of the stratosphere. While the physics of the troposphere of planets is well understood, the stratosphere has been only imperfectly worked out. Whereas energy transport below the tropopause is by convection, above the tropopause it is by radiation. But the radiation transfer is not that in a uniformly grey atmosphere. The molecular absorption is in the thousands of lines in the absorption bands of water and carbon dioxide in the terrestrial planets and of methane and ammonia in the major planets. The absorption coefficient changes rapidly with respect to both wavelength and altitude. A full treatment of the problem requires numerical integration with sufficiently small intervals of these parameters. It has been

estimated (Goody, 1954) that about 10^7 columns would be required. Furthermore, the absorption coefficient is only imperfectly known for each of the lines in the earth's atmosphere. Obviously simplification by statistical treatment of the problem is required and progress is being made in this direction.

The absorption bands fall into two general types which result from the form of the molecule. Linear and symmetric-top molecules have their lines at more or less uniform spacing and intensity. W. M. Elsasser has worked out a model based on this idea and which predicts the amount of absorption. On the other hand, asymmetric-top molecules, such as water have their lines more or less at random. Goody has worked out a model for lines of random spacing and intensity. But ozone, though an assymmetric molecule, is only slightly so and has its lines arranged in the direction of uniform spacing. A model predicting the amount of absorption with an adjustable randomness parameter has even been given.

The line shape must be taken into account and in the above models the Lorentz collision broadened line is used which is given by

$$k_\nu = (s/\pi)\alpha_L/[(\nu - \nu_0)^2 + \alpha_L^2]$$

where k_ν is the absorption coefficient and s is the total line strength $= \int k_\nu d\nu$. The line half width α_L is proportional to the total pressure which, of course, varies exponentially with altitude. Somewhere in the stratosphere the collision width becomes substantially less than the Doppler width and the line shape is a hybrid of the two shapes. The models given above break down at this point. Furthermore the Lorentz shape is only approximate, especially in the far wings.

One would like to have a theory which gives the altitude and temperature of the tropopause with fair accuracy and as input data, the composition of the atmosphere and the surface temperature. The composition should be translatable into a few parameters such as line width, randomness parameter, and pressure.

Not having this theory, we shall, because of the insight into the problem, summarize the theory of Emden (1913). Emden assumed for the earth that it had a grey atmosphere between 4μ and $\infty\mu$ which absorbs 90% of the outward flux, water vapor produces this absorption and has a scale height of 2 km, all of the solar radiation reaches the ground, and the low temperature radiation is in two antiparallel vertical streams. He first assumes that the entire atmosphere is in radiative equilibrium. The temperature very high in the atmosphere becomes $(\frac{1}{2})^{\frac{1}{4}}$(surface temperature) $= 202°$K. Within 3 or 4 km of the ground the lapse rate is much greater than adiabatic and there is a 19° discontinuity of temperature between the ground and the base of the atmosphere. Convection will set in and straighten the curve to the adiabatic rate until intersection with the radiative-equilibrium curve.

The stratosphere will be terminated at the top by absorption of solar radiation resulting in the dissociation of molecular constituents, and resultant heating of the atmosphere. This occurs on the earth in the Schumann continuum of O_2, giving rise eventually to ozone, which in turn absorbs in its Hartley continuum. If oxygen is nearly absent on Venus and Mars, carbon

dioxide will operate in the same manner, absorbing beyond its dissociation limit at 1690 Å and producing CO + O.

We should now turn to the empirical results for the planets, but since Mars and Venus are best observed, and consequently best understood, we will place most of our attention on these.

The atmospheric pressure at the surface of Mars has been derived by several methods. Only the most reliable, which is the result of polarimetric measurements by A. Dollfus (1951a), is presented here.

Dollfus made extensive measurements of the polarization of bright equatorial regions of Mars through orange and green filters and at points near the center of the disk and near to the edge of the disk. The light arises from reflection from the surface and Rayleigh scattering by the atmosphere. The brightness of the atmosphere is $B_a(1 + \cos^2 i/2\cos\theta)(\lambda_0/\lambda)^4$ and the brightness of the surface is $B_s\phi(i)\Psi(\theta)$ where i is the phase angle, θ is the planetocentric angle of the observed area from the center of the disk, and B_a and B_s are the brightnesses of the atmosphere and surface at the center of the disk at opposition and at λ_0. The polarization of the surface is given by $P_s(i) + \Delta(\theta)$ and the polarization of the atmosphere is $\sin^2 i/(1 + \cos^2 i)$. The total observed polarization if $B_a << B_s$ is

$$P(i,\lambda,\theta) = P_s(i) + \Delta(\theta) + (B_a/B_s)(\lambda_0/\lambda)^4(\sin^2 i/2\phi(i)\Psi(\theta)\cos\theta).$$

Dollfus applied this formula to his observations in three different ways:

1) by comparing the polarization at $\theta = 60°$ to that at $\theta = 0$ for different phase angles. This gave $B_a/B_s = 0.040$.

2) By using different wavelengths, 5100 A and 6200 A. This application yielded $B_a/B_s = 0.035$.

3) By comparing the polarizations of surface areas of different brightnesses between center and limb. This gave $B_a/B_s = 0.034$.

He also measured the total polarization of the disk and by using different wavelengths he was able to derive $B_a/B_s = 0.037$.

From the magnitude of Mars at mean opposition, B_s can be found and therefore B_a, which can then be converted to the total amount of gas by assuming that the scattering is the same as for air. The atmospheric pressure thus found is 83 mb = 62 mm of Hg.

The polarimetric method has not been applied to other planets because of the presence of clouds and consequently large and variable polarizations prevailing. The atmospheric pressure on Venus can be found by several methods which will be described below.

The composition of the atmospheres of most of the planets has been fairly well established. The table gives the spectroscopically observed constituents and the amounts. The amounts, however, represent only that part above or perhaps slightly below the cloud layer or surface if that is seen directly.

The determination of the amount of gas from the intensity of its absorption band has considerable theoretical difficulties. The shapes of the lines is given sufficiently accurately by the Lorentz formula or the Doppler line shape provided a hybrid shape is not involved. W. M. Elsasser's model, which may be used for diatomic, linear, and spherical top molecules, has an infinite

array of uniformly spaced lines of spacing d, Lorentz half line width α, line strength S, and optical thickness x. Elsasser found two simple formula which apply in certain useful limiting cases:

$$1)\ T = 1 - \phi(\pi S\alpha x/d^2)^{1/2} \quad \text{if} \quad 2\pi\alpha/d << 1$$

where ϕ is the probability function, i.e., $\phi(y) = 2/(\pi)^{1/2}\int_0^y e^{-t^2}dt$

$$2)\ T = e^{-(Sx/d)\tanh(2\pi\alpha/d)} \quad \text{if} \quad 2\pi\alpha/d >> 1.$$

Equation 1) is probably the most useful to us. For small x, it may be simplified further to

$$T = 1 - 2(S\alpha x/d^2)^{1/2}$$

which is the well-known "square-root law". The line width, α, is generally found to be proportional to the total pressure and hence the absorption is proportional to the square root of the pressure and the square root of the amount of absorbing gas. The difficulty in applying this expression to planetary atmospheres is two-fold. One often does not know the atmospheric pressure sufficiently well. Another difficulty is that the pressure is not constant through the atmosphere. The pressure variation can, however, be allowed for.

The determination of the amount of a gas is also straight forward if very weak lines can be measured under high resolution. The usual exponential law of transmission applies and Beer's law is obeyed, i.e., the total pressure is irrelevant. Because the exponent is much less than one, the absorption varies linearly with the amount of gas. As a consequence much emphasis is to be placed on measurements made in bands where only weak absorption occurs.

Another important case where the absorption is independent of pressure occurs when Doppler broadening dominates collision broadening. Lorentz broadening in terms of frequency units is nearly the same for all absorption bands of a gas and since Doppler broadening is proportional to frequency, the higher the frequency—the greater the ratio of Doppler to collision broadening. As a rule the absorption of bands lying in the visible is independent of pressure if they are not too strongly absorbing. With these cautions in mind the atmospheric compositions are presented.

Venus Adams and Dunham (1932) found bands at 7820, 7882, and 8689 A which they concluded were CO_2 from the observed rotational structure. These bands have been reproduced in the laboratory by Adel and Slipher (1934) and by Dunham, and by Herzberg (1949). The most reliable amount is 10^5cm-atmo. Kuiper and Chamberlain (1955) derived 285°K from the rotational structure. The partial pressure of CO_2 at this level will be 170 mb. If we assume the surface is 580°K and $\gamma = 1.3$ for CO_2 the pressure of CO_2 at the surface is 3.4×10^3 mb. We expect that about the same amount of N_2 will be on Venus as on the earth and allowing for the lower gravity the total pressure is about 4000 mb.

The amount of CO_2 in the terrestrial atmosphere has been shown by Urey (1952) to be controlled by reactions of the type

$$MgSiO_3 + CO_2 \rightarrow MgCO_3 + SiO_2.$$

At 300°K the equilibrium pressure of CO_2 is 10^{-5} atmo. The pressure of CO_2 on the earth is 3×10^{-4} atmo, i.e., within a factor of 30 of the above. The low amount of CO_2 is maintained by erosion which brings fresh rock surfaces into contact with the CO_2. Water acts as a catalyst as well as producing the erosion. High temperatures reverse the reaction, and consequently volcanic activity releases CO_2. From the large amount of CO_2 present on Venus and probably an amount of volcanic activity similar to the earth's we can expect that there is little erosion.

Strong (private communication) from the balloon observations of Ross and Moore has determined 1.9×10^{-2} g/cm^2 of H_2O, which is presumably above the cloud layer. There is the slight possibility that terrestrial water contributed significantly to the measurement. The author has recently obtained spectra of Venus between 1 and 3.5μ. These show a sharp drop relative to the sun's spectrum beyond 2.7μ and it is believed that the explanation lies in absorption at 3.1μ by ice crystals in the clouds. Lyot (1927) determined that the polarization curve of the clouds resembled that of water drops of 2.5μ radius. Perhaps better agreement might be obtained with ice crystals of this size. Deirmendjian, Classen, and Xiezee (1961) give calculations of Mie scattering by absorbing spheres. For spheres of 2.5μ radius and the absorption coefficient of ice at 3.1μ it appears that ice will explain the observed spectrum. In view of these supporting observations the balloon results seem reasonable.

The carbon monoxide band at 2.35μ has been observed by the author. The amount is about 4 cm-atmo and it is probably found only above the cloud layer where it is produced by photodissociation of CO_2. Therefore the observed amount is the total amount in the whole atmosphere. Kozyrev (1954) found two absorption features at 4120 and 4378 A. These have been identified by Urey as belonging to CO^+. Kozyrev's observations, however, are in doubt and were only partially confirmed by Newkirk (1959).

Tests for other gases have been negative with the following results: $O_2 <$ 500 cm-atmo, $N_2O < 100$, $CH_4 < 20$, $NH_3 < 4$, and $O_3 < 0.05$ (Kuiper, 1952).

It has been suggested by several persons (Sinton, 1953; Kuiper, 1956) that carbon suboxide, C_3O_2, might be present. It would be produced by the following reaction: $2CO + CO_2 + h\nu = C_3O_2 + O_2$. Its characteristic absorption bands at $\lambda < 3350$ A have not been found despite searches by Kuiper and also by Sinton. Carbon suboxide, however, polymerizes to form a yellowish powder which might explain the yellowish color of Venus. A band found by Sinton and Strong (1960a) at 11.2μ in the spectrum of Venus might be produced by the polymer, but laboratory comparisons are ambiguous.

The stable atmospheric gases will have volume percentages close to: CO_2 — 75%, N_2 — 25%.

Mars The only gas definitely detected in the atmosphere of Mars thus far is carbon dioxide. Kuiper (1952) found CO_2 bands at 1.57, 1.60, 1.96, 2.01 and 2.06μ. Sinton and Strong (1960b) also observed bands at 9.4, 10.4, and 12.6μ. Kuiper found that the strength of the 1.57 and 1.60μ bands when observed with Mars near the zenith was equivalent to the terrestrial strength of the bands when the moon was observed at 10 air masses. Grandjean and Goody (1955) derived an amount of 3600 cm-atmo from Kuiper's data.

The Martian CO_2 pressure is far from that given by the Urey equilibrium, and as a result we expect that there is relatively little erosion there. Certainly there is rarely liquid water available.

Kuiper (1952) also found that the polar cap on Mars showed a sudden decrease in reflectivity beyond 1.5μ as does frost. It is therefore quite clear that there is some moisture in the air. It, however, has never definitely been found spectroscopically. V. M. Slipher (1924, Very, 1909) in the winter of 1908, when he knew that the terrestrial water above Flagstaff was extremely small, obtained a spectrum of Mars which showed the band more intense than in spectra of the moon. It is now known that on rare occasions, such as that employed by Slipher, the amount of water above Flagstaff may be 0.02 g/cm^2. More recent searches with similar sensitivity have failed to find Martian water (Dunham, 1952; Kiess, 1957). The amount of water vapor in the Martian atmosphere is undoubtedly variable; the water is locked in the polar caps in winter and released in summer. This fact may explain the detection by Slipher and the failure to do so by Adams and Dunham.

The amount of water vapor may, however, be estimated from the formation of white obscuring haze on the morning limb of the disk. The temperature at this point has been estimated at —70°C (Sinton and Strong, 1960b). The frost point of the water vapor is obviously above —70°. A lower limit is therefore a vapor pressure of 0.002 mm of mercury giving a total water content of 0.004 g/cm^2 (Urey, 1959). An amount twice this may have been detectable in Slipher's spectra.

Tests made for oxygen have also been largely negative and have set the total amount at less than 240 cm-atmo. In addition tests by Kuiper for ozone have set this gas at less than 0.05 cm-atmo.

It is supposed that the total atmospheric pressure of 83 mb is furnished mostly by nitrogen. If this be so, then the amount of nitrogen is 1.8 x 10^5 cm-atmo. Argon is supposed also to be present in the amount of about 200 cm-atmo.

In the absence of oxygen, CO_2 should be photodissociated as it is on Venus and an amount of CO should be present of the same order as that on Venus, i.e., about 4 cm-atmo. The volume percentages of the stable gases are close to: N_2, 97.3; CO_2, 1.6; and A, 1.1.

The Major Planets In contrast to Mars, Venus, and the earth, which have oxidized atmospheres, the major planets have highly reduced atmospheres. It is not my purpose to go into detail concerning these planets here. Urey (1959) and Kuiper (1952) have drawn up models for the atmospheres of

these planets. I feel that at the moment it is quite difficult to construct reliable models. The assumption is often used for Jupiter and Saturn that the solar abundances of H, He, and C are maintained. But these planets have perhaps 10% to 1% of their original masses and fractionation of the solar composition is bound to occur. Four lines of the 3 - 0 band of hydrogen were recently found by Kiess in the spectrum of Jupiter and from intensity measurements on these we may have a much better model.

Atmospheres II

A fairly definite model of the atmosphere of Venus can now be constructed. It is based largely on observational evidence with theoretical laws to tie the various observed parameters together. In addition there are in some respects over determinations of the model, and these check with each other. For example, the pressure at three different levels in the atmosphere has been determined and these more or less agree using the temperature model for the scale height. In drawing up the model it is embarrassing, however, that we still do not know the rotation period.

There are two recent observations that have helped immeasurably in piecing together the observational data. These are the radio observations of thermal emission, which presumably arises from the surface and they give its temperature, and the photometric observations that were made of the occultation of Regulus. The radio temperature of 580°K as first observed by Mayer, McCullough and Sloanaker (1958) is now well established and this value should be close to the temperature at the base of the atmosphere. The occultation observations, reported by de Vaucouleurs and Menzel (1960) give the temperature near the "top" of the atmosphere, and at a definitely known altitude.

Temperatures of intermediate regions are supplied by the 8 - 13μ measurements (Pettit and Nicholson, 1955; Sinton and Strong, 1960a) and by the temperature of the near IR reflecting layer as determined from the intensity distribution within the CO_2 bands (Adel, 1937; Chamberlain and Kuiper, 1956).

The author (unpublished) determined the temperature of the dark hemisphere at the 1961 inferior conjunction at 3.75μ and obtained 236°K. As mentioned before, the radiation beyond 3μ appears to be stopped by ice crystals and we can assume that this temperature applies to the top of the visible clouds. Since 8 — 13-μ radiation would also be stopped at this point and the above temperature is identical to the 8 — 13-μ temperature, it is assumed that the latter temperature also refers to the top of the visible clouds. Sagan (1961) has pointed out that the temperature determined by Chamberlain and Kuiper does not refer to the top of the clouds but to an undetermined depth below the clouds.

It is fortunate that we have two temperatures corresponding to definite levels in the atmosphere. The other temperature is that obtained from the occultation of Regulus. De Vaucouleurs and Menzel found 297°K, a lapse rate of 3°/km, and an atmospheric pressure of 2.6×10^{-6} atmospheres. The

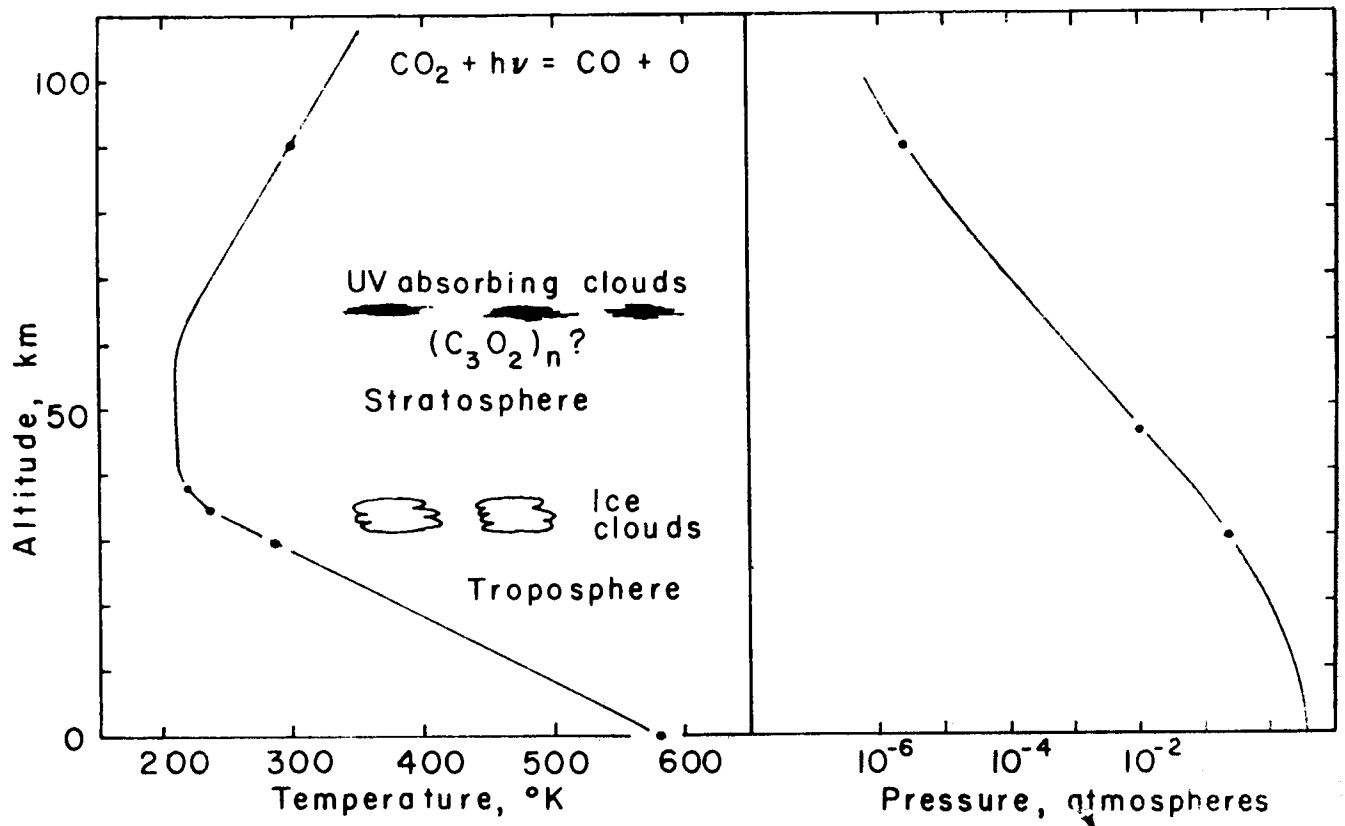

$CO_2 + h\nu = CO + O$
UV absorbing clouds
$(C_3O_2)_n$?
Stratosphere
Ice clouds
Troposphere
Altitude, km
100
50
0
200
300
400
500
600
Temperature, °K
10^{-6}
10^{-4}
10^{-2}
Pressure, atmospheres

Fig. 14

level is obtained from the duration of the occultation. The radius, which was obtained in this way, is 6125 ± 2 km. The optical radius, representing the top of the cloud level, is 6095 ± 7 km. The occultation temperature and pressure accordingly are at a level 70 ± 8 km. above the cloud level. The chief error in the altitude difference is due to the error in the radius at the cloud level, and unfortunately amounts to more than a scale height.

Assuming, as before, the adiabatic lapse rate of −10°/km, one finds the depth of the surface below the cloud level, i.e., (580 − 235)/10 = 34.5 km. The depth corresponding to the temperature determined from the CO_2 bands is (285 − 235)/10 = 5 km. One might easily argue that since it is now fairly certain that there is water in the atmosphere, the lapse rate should be less than the adiabatic lapse rate. Accordingly, these depths may require revision.

Sinton and Strong (1960a) found a limb darkening in the 8 — 13-μ region which was given by $I = I_0 \cos^{1/2}\theta$, θ being, as before, the planetocentric angle from the center of the disk. J. I. F King (1960) has shown that this leads to a negative lapse rate of about 6°/km. The gist of the argument runs as follows. The observed distribution of radiation over the disk will be given by

$$I(\mu) = \int_0^\infty B(\tau)e^{-\tau/\mu}d\tau/\mu = I(1)\mu^{1/2}$$

where μ is written for cos θ, and τ is the optical depth and B is the emission as a function of depth. Inverting the Laplace transform, gives $B(\tau) = 2\pi^{-1/2} I(1)\tau^{1/2}$. Since B and I are proportional to T^4 we can express the above as $T(\tau) \propto T(1)\tau^{1/8}$. The lapse rate is

$$-dT/dz = -(dT/d\tau)(d\tau/dz), \quad \text{and} \quad d\tau/dz = -\alpha\rho$$

From the gas law $\rho = pm/kT$, and since $p = mg = \tau g/\alpha$,

$$\rho = -(mg/k\alpha)(\tau/T),$$

Therefore $d\tau/dz = (mg/k)(\tau/T)$. Substituting we have

$$-dT/dz = (mg/k)(dT/d\tau)(\tau/T = (mg/k)(d \log T/d \log \tau).$$

But since $\log T = C + (1/8) \log \tau$, the lapse rate is $-dT/dz = (1/8)mg/k$.

The adiabatic lapse rate is $[(\gamma - 1)/\gamma]mg/k$, and therefore the observed lapse rate is $(1/8)\gamma/(\gamma - 1)$ of the adiabatic. For CO_2 the observed lapse rate is 6°/km. The lowest temperature observed near to the limb is close to 218°K and the extent of the observed lapse rate is from this point to 235°K.

With our temperature diagram which is now constructed (see Fig. 14) one can see if the two observed pressures (2.6×10^{-6} atmo at 70 km above the cloud level, and 0.2 atmo at 5 km below the top of the cloud level) are consistent with the scale height derived from the temperatures. Using a different median scale height kT/mg for each 5 km interval and allowing an adjustable thickness for the isothermal region, we find that 60 km are

required for the pressure to decrease from 0.2 atmo to 2.6×10^{-6} atmo. The difference of 15 km can be attributed to inaccurate diameter measurements, and we accept the altitude difference of 60 km as more nearly correct.

There is one other pressure determination which is available. Venus occasionally transits the sun. The last ones were in 1874 and 1882 and the next ones are not until 2004 and 2012. A curious phenomenon is noted just as the disk of Venus is encroaching or leaving the solar disk. The planet's limb that is away from the sun becomes illuminated. It has long been realized that the cause is light which is refracted through the planet's atmosphere; in the case of Venus the total deviation is about 88″. The observed light just skims the top of the cloud level. The theory of the deviation is well known (c.f. Fabry, 1929) and the deviation is given by

$$\omega = \epsilon(2\pi R/H)^{1/2}$$

where ϵ is the difference between the index of refraction and unity at the lowest level pentrated in the atmosphere, H is the scale height, and R is the radius of the planet. For Venus we derive $\epsilon = 4.89 \times 10^{-6}$. For CO_2 at NTP $\epsilon = 4.5 \times 10^{-4}$. The quantity ϵ is proportional to pressure and inversely proportional to temperature. At 235°K the required pressure is 1.0×10^{-2} atmo. The height to which this pressure applies, however, is somewhat uncertain. During a transit one will see a band of light of thickness corresponding to the altitude of the level above the cloud level having this pressure. If this altitude is only 1 km, then its apparent thickness to an earth-bound observer would be only 0.005″ and its true surface brightness would be nearly the same as the solar disk since the reduced thickness of gas penetrated is only 4 km at NTP. But because the terrestrial seeing would blur the image over about 3″, its apparent brightness would be 0.17% of the sun and doubtlessly not be seen so close to the sun's limb. On the other hand if the altitude is 12 km above the clouds, the apparent brightness would be 2% of the sun and could probably just be seen. We will adopt 12 km above the cloud layer for the pressure of 0.01 atmo. It agrees well with the adopted pressure curve.

There are a few other facts to be noted about the model. The reduced thickness of gas above the occultation level at 90 km is 1.7 cm at NTP. The absorption coefficient of CO_2 just beyond the 1690 A edge of the continuous absorption is 1 cm^{-1}. The dissociation-effective ultraviolet light will penetrate about to the 90-km level. Carbon monoxide, formed here, will diffuse downward, but will, as pressure increases, be oxidized to CO_2 as the frequency of 3-body collisions increases. Maximum CO amount probably occurs in the 50 to 90-km region, and as noted, the total amount is 3 to 4 cm.

Carbon suboxide may be formed at a high level, and the polymerized molecule may explain the dark ultraviolet clouds that are seen.

The clouds observed in the visible appear to be fairly definitely ice crystals and the amount of water observed by Strong is approximately that expected of saturation at 235°K.

Though the model presented here appears to hang together well and satisfies all known observations, it hinges on the assumption that the observed

microwave radiation arises from the surface. On the other hand it is not yet definitely established that it can not arise from an ionosphere. If an ionosphere is the cause of the observed emission, it would have to possess about 10^3 times the electron density of the terrestrial ionosphere. This level of ionization can not be supplied by solar ultraviolet light, but if Venus has a very small magnetic field the amount of ionization could be supplied by the solar wind.

The information from which we can build an atmospheric model for Mars is much inferior to that which we have for Venus. Perhaps we do know with higher accuracy the atmospheric pressure at the surface (the previously mentioned 83 mb). But the only observed temperature available is that of the surface. This has been measured by Coblentz and Lampland (1927), Pettit and Nicholson (1924), and Sinton and Strong (1960b). The maximum surface temperature, as we have seen, is close to 25°C. For the atmospheric temperature close to the ground we will assume between 0 and 10°C at the maximum. The previously adopted composition (N_2, 97.3%; CO_2, 1.6%; and A, 1.1%) gives a γ of 1.4 and an adiabatic lapse rate of 3.7°/km which we will assume. The greenhouse effect will be much reduced on Mars relative to the earth owing to the smaller water vapor content, but in part this lack will be replaced by the greater CO_2 abundance. One might assume for the stratospheric temperature that which is given by Emden's theory, i.e., $(\frac{1}{2})^{\frac{1}{4}}$ of the surface temperature, or 238°K. The height of the tropopause becomes $(283 - 238)/3.7 = 12$ km. This temperature is obviously too high, for condensation occurs in the Martian atmosphere and with the spectroscopically determined upper limit to the amount of water, there is not enough water for condensation at this temperature. Hess (1958) has shown that the scattering properties of the blue haze layer lead to a probable surface frost temperature of —90°C and a total water content of 0.002 mm. For condensation to occur the stratosphere temperature would have to be about 170°K and the tropopause altitude would have to be 30 km. Other atmospheric models have been given—by de Vaucouleurs (113°K, 45 km)(1954), Kuiper (182°K, 9 km)(1952), and Urey (180°K, 30 km) (1959). For our model we will use the earth as a guide and make the tropopause temperature

$$T_{t\,♂} = T_{s\,♂}\, T_{t⊕}/T_{s⊕}$$

where t and s refer to tropopause and surface. If $T_{s\,♂} = 273$, $T_{s⊕} = 290$, and $T_{t⊕} = 218$, then $T_{t\,♂} = 205$°K. Condensation can occur at this temperature if the surface frost temperature is —60°C and the total atmospheric water is but 0.14 mm, an amount which would not have been detected spectroscopically. The height of the tropopause is 18 km.

Earlier we derived the amount of water from the assumption that condensation occurs on the surface in the morning. This gave a frost point of — 70°C and a total water content of 0.004 mm. A temperature inversion probably occurs at night making the air near to the ground the coldest in the atmosphere. We can feel fairly confident that the observed condensation is actually on the ground.

In the absence of appreciable oxygen, CO_2 will be disassociated in the high atmosphere of Mars as it is on Venus. About 1 cm of gas will be penetrated by $\lambda < 1650$ A. The scale height with $T = 205°K$, $\bar{m} = 28$, and $g = 370$ is 16.3 km. The pressure at the tropopause is $p = p_0(T/T_0)^{\gamma/(\gamma-1)} = 0.8$ mb with $p_0 = 83$ mb and $\gamma = 1.4$. The CO_2 penetrated will exert a partial pressure of $p = wLmg$, which is 0.725 dynes/cm^2 $= 0.725 \times 10^{-3}$ mb. The altitude of this level above the tropopause is given by $0.029/0.8 = e^{-h/16.3}$ or $h = 54$ km. The height above the surface is 72 km. The total amount of CO formed should be about the same as on Venus. The amount of solar energy absorbed to this level will be about ¼ of that absorbed to the same level on Venus because of the greater distance from the sun.

I feel that the blue haze layer is caused primarily by scattering. E. C. Slipher's blue photographs show limb brightening—not darkening, as assumed by Opik (1960) in his recent discussion of the absorbing properties of the layer. Both absorption and scattering may, however, be important, because the blue photographs sometimes show dark clouds against a uniform background. By some unknown mechanism the blue haze has a tendency to evaporate near opposition with the earth. This has definitely been established and will be published shortly by Slipher.

The moon probably has a very tenuous atmosphere that is not detectable by the usual methods. Elsmore (1957) determined from the times of occultation of a radio source that the electron density was 10^3/cm^3 in the vicinity of the moon. Since gas would be partly ionized by solar ultraviolet and by the solar wind, this figure is close to the upper limit to the density of the atmosphere.

Recently, Watson, Murray, and Brown (1961) have shown that water is one of the most stable volatile substances on the lunar surface. Their argument runs as follows. The escape of a volatile is limited by its evaporation rate from a cold-trap area of the moon (the permanently shaded regions near the pole of the moon). The temperature of these regions is estimated at 120°K or less. The rate of loss of water from the moon's surface is given by,

$$\dot{m}_s = KA(\dot{E} - \dot{C})$$

where K is the fraction of permanently shaded areas (estimated at 5×10^{-3}), A is the area of the lunar surface, and $\dot{E}$ and $\dot{C}$ are the evaporation and condensation rates. The evaporation rate depends on the vapor pressure of the material at the cold trap temperature, and is given by

$$\dot{E} = p(m/2\pi RT)^{1/2}$$

where p is the vapor pressure of the volatile of molecular weight m at temperature T.

The mean free path for a molecule evaporating from the cold trap is much greater than the radius of the moon, and since its velocity corresponds to a very low temperature, it can not escape the moon but must fall back following a ballistic trajectory without suffering collisions. It, of course,

has only one chance in 2000 of landing in a cold trap, but it reevaporates and eventually finds its way back to a cold trap again, unless it can escape either by being "blown" by the solar wind or by being disassociated by ultraviolet radiation. If the chances are high that it escapes, the condensation rate is negligible and the escape rate is the rate of evaporation from the cold trap. On the other hand if the chance of escape is small compared to the evaporation rate, the condensation rate will almost equal the evaporation rate. In this case the moon has an atmosphere whose pressure is the vapor pressure corresponding to the cold trap temperature. For water vapor the computed escape probability for the two methods together is only 4×10^{-3}. Therefore the moon will have a water vapor atmosphere at the vapor pressure of the cold trap. If the cold trap temperature is 120°K the density of H_2O molecules is $3.5 \times 10^4/cm^3$. The ionization produced by the solar wind means that 4×10^{-3} of these are ionized, or about 10^2 ions/cm^3. This result is in fair agreement with the density of 10^3 found by Elsmore. It should be noted that 120°K is an upper limit to the temperature of the cold traps. The actual temperature is probably closer to 100°K.

RADAR METHODS OF STUDYING THE EARTH'S IONOSPHERE

JOHN V. EVANS
*Lincoln Laboratory**
Massachusetts Institute of Technology

I. Introduction

In 1882, Balfour Stewart[83] proposed that there were electric currents flowing in the Earth's upper atmosphere. He made this suggestion to explain the small regular variations which had been noted in the strength of the Earth's magnetic field. This idea, that the Earth has an electrically charged outer atmosphere, was revived in 1902 when both Heaviside[40] and Kennelly[48] independently proposed that such a layer might exist, and be a reflector of radio waves. In this way they sought to explain Marconi's observation (in 1901) that radio signals could be transmitted across the Atlantic Ocean.

Direct experimental confirmation of the existence of such a region was not forthcoming until 1925 when Appleton and Barnett[3] observed at a distant receiving site the signals from a CW transmitter when it was slowly frequency modulated. They found that the received signals showed systematic fading, which indicated that two signals were arriving by different paths and interfering at the receiver. One signal was taken to be the direct signal (or ground wave) whilst the other signal they concluded had been reflected from a region in the atmosphere at a height of 60 miles. In 1926, Breit and Tuve[16] developed their pulse sounding technique of studying these layers, which at that time were known as Kennelly-Heaviside or sometimes Appleton layers. The term "ionosphere" was introduced later by Watson Watt.

The pulse sounding technique came to dominate the study of ionosphere for the next twenty-five years. Vertical incidence sounders were established in many countries, and this work was intensified considerably during the years of the Second World War. Thus by about 1945, the morphology of the ionosphere was well known. It had been established that there are normally three distinct layers which were called the D, E and F layers after a notation introduced by Appleton. The F region was also known to split into two layers (the F_1 and F_2) during the daytime at most latitudes. A great deal of theoretical work had been directed towards the reduction of ionospheric soundings, to yield the true distribution of electrons with height. This work is laborious as it can best be approached by a process of successive approximations. Thus the reduction of such records is not usually attempted nowadays without the aid of a digital computer.

*Operated with support from the U. S. Army, Navy and Air Force.

Other theoretical work had been concerned with the shape and variation of these layers. The D, E and F_1 layers were found to vary in density throughout the day and seasons in a manner predicted in a classic paper by Chapman[19] in 1931. However, the F_2 region was shown to be distinctly anomalous. Because an ionospheric sounder can not yield information about the distribution of electrons above the peak of the F_2 layer, there was considerable speculation about this inaccessible region. It was suggested that perhaps the total number of electrons in the F_2 region might vary according to Chapman's theory, but the shape of the region might be altered perhaps by electrodynamic forces to give rise to an anomalous behavior for the maximum electron density.

During the last five years, four new methods of studying the ionosphere have been developed and these have provided information about the region above the F_2 peak. These are:

A. Rocket probes
B. Moon-reflected signals
C. Signals from artificial earth satellites
D. Incoherent "backscatter" echoes

Although rocket probes penetrated into the upper F layer as early as 1949,[7] these measurements have only recently reached a degree of precision comparable with that of ionospheric sounders. Further, because the upper part of the F region has a greater extent than the lower part, rocket flights are required to reach an altitude of about 600 kms in order to give a useful result concerning the distribution of electrons. Thus the first results which provided some ideas concerning the true extent of the ionosphere were those obtained using moon-reflected signals.[26] In these experiments the rotation of a plane polarized wave caused by the Faraday effect in the Earth's ionosphere was measured. The moon, which fortunately does not depolarize the signal greatly, is used as a convenient reflector and the waves then suffer the same additional rotation on their return path. Following close on these measurements (which were first made in 1955) similar measurements were made using the signals of artificial earth satellites. Both these techniques are currently being exploited to yield accurate values of the total electron content of the ionosphere When examined jointly with the appropriate vertical incidence ionospheric soundings, they can yield a good deal of information concerning the gross shape of the ionosphere.

The fourth technique was proposed by Gordon[34] in 1958. In this experimental method a very high power short-pulse radar is directed upwards at the ionospheric layers and weak incoherent echoes from the individual electrons are obtained. These reflections occur despite the use of an exploring radiowave frequency many times higher than the critical or plasma frequency for the region. Such echoes are not predicted by magneto-ionic theory for the propagation of radio waves in the ionosphere, which neglects the discrete nature of the electrons. Although only three years have elapsed since Gordon's proposal, this method has already been put to the test, and a considerable body of experimental results have been obtained.

In this lecture I propose to review these new results, obtained by means of radar reflections from the moon, and by the radar incoherent scatter method. I shall also consider how well present day theories for the F region are in accord with these latest observations.

II. Moon-Echo Observations

The earliest workers to obtain radar reflections from the moon were members of the U. S. Army Signals Laboratory, under the direction of Lt. Col. J. H. DeWitt. They succeeded in detecting echoes first on January 10, 1946 (Webb[88], Mofensen[59]), but a full account of this work was not published until 1949 (DeWitt and Stodola[21]). This group found that the echo amplitudes showed marked fading, and on some days the echoes were absent altogether. Later in 1949 workers at the Australian Scientific and Industrial Research Organization (Kerr, Shain and Higgins[49]) commenced a series of observations using short-wave broadcast transmitters and conventional communications-type receivers. Despite the fact that the interpretation of their results was complicated by ionospheric effects (ionospheric refraction and possibly absorption) Kerr and Shain[50] were able to distinguish two forms of fading. The first having a period of a few seconds was attributed to the rocking or libration of the moon.* In support of this, Kerr and Shain were able to show that the fading rate of the echoes was related to the rate of libration in almost exactly the manner that would be expected if the moon were to behave as a uniformly bright scatterer (a conclusion not borne out by later workers: Evans[25], Trexler[86]). The second fading mechanism appeared to be responsible for variations in the mean intensity of the echoes, and this had a period of the order of minutes. They suggested (wrongly, as it was later discovered) that this long-period fading might be caused by irregularities in the earth's ionosphere which serve to defocus the aerial beam. Such irregularities are observed to cause fluctuations in the intensity of radio stars, but usually with a period much shorter than several minutes.

The origin of the long-period fading was not solved until 1954 when Murray and Hargreaves[62], who were working at the Jodrell Bank Experimental Station in England, were able to show that it was caused by the Faraday effect in the earth's ionosphere. This effect is brought about by the presence of the earth's magnetic field, which causes the ionosphere to have two refractive indices for waves which are propagated along or nearly along the lines of the earth's magnetic field. As a result a plane wave is propagated as two circularly polarized waves having opposite senses of rotation and different phase velocities. When the wave emerges from the ionosphere, the two components recombine to form a plane wave which has been rotated relative to its original position. Fortunately the moon causes only about 10 percent depolarization (Browne *et al.*[17]) and hence a wave, which is still

*The moon does not keep exactly the same face turned towards an observer on the earth—largely because of the parallactic shift caused by the observer's own motion. This apparent rotation of the moon is called libration.

largely plane, is reflected back towards the earth. The wave then suffers the same amount of rotation on retraversing the ionosphere—in a sense which makes it add to the rotation introduced by the first passage. The total number of rotations (2π radians) Ω has been shown to be[17]

$$\Omega = \frac{1}{2\pi^2} \frac{e^3c}{m^2f^2} \int_0^R N_r B_r \cos \theta_r \, dr \qquad \text{rotations for the two-way path} \qquad (1)$$

where e and m are the charge and mass of an electron, c is the velocity of light and f the frequency of the radio waves. N_r is the electron density, B_r is the intensity of the earth's magnetic field, θ_r is the angle between the ray path and the magnetic field lines and r is the distance measured along the line-of-sight to the moon (at a distance R). If the electron density N is written as a function of height h above the surface of the earth, instead of range r, the expression becomes

$$\Omega = \frac{1}{2\pi^2} \frac{e^3c}{m^2f^2} \int_0^H N_h B_h \cos \theta_h \sec \chi_h \, dh \qquad \text{rotations for the two-way path} \qquad (2)$$

where χ_h is the angle between the ray and the zenith. Browne and Evans[18] (and later Bauer and Daniels[6]) have shown that provided there are no more than about 10^3 electrons cm^{-3} in the space beyond the earth and the moon, most of the rotation Ω occurs within the first 1000 kms height above the earth. Thus (2) can be written

$$\Omega = \frac{7.53 \times 10^3}{f^2} \cdot \overline{B \cos \theta \sec \chi} \int_0^{1000} N dh \qquad \text{rotations} \qquad (3)$$

The appropriate value of $\overline{B \cos \theta \sec \chi}$ has been discussed by Yeh and Gonzalez[92] and in practice is close to the actual value computed for a height in the ionosphere lying between 300 and 400 kms. Equation (3) is accurate to within about 2 percent provided that a radio wave frequency f of about 100 Mcps is used. For such a frequency, the radio wave may propagate at angles θ up to about 89° to the field before the assumptions made in the derivation of (1) break down. The integral $\int_0^{1000} Ndh$ is the total electron content of the ionosphere and if this is written n_t then it follows from (3) that

$$\Omega \propto n_t/f^2 \qquad (4)$$

Thus by measuring Ω we can determine n_t. Unfortunately, unless a very high frequency f is chosen, the value of Ω will be ambiguous as it will not be known in what quadrant the plane of polarization of the reflected waves lie, nor how many complete rotations the waves have undergone. To resolve this ambiguity, Browne *et al.*[17] used two close-spaced frequencies

separated by Δf. The difference in the amount of rotation $\Delta\Omega$ between the frequencies will then be

$$\Delta\Omega = \frac{2\Delta f}{f} \qquad (5)$$

and by measuring $\Delta\Omega$ the total rotation Ω and thence the total electron content n_t can be determined.

The first systematic measurements of n_t by this method were made in 1955 and 1956 at the Jodrell Bank Experimental Station by Evans[24, 26]. The principal results of this work were

(a) The ionosphere does not have a simple parabolic form (see Ratcliffe[74]) because almost three times as many electrons n_a exist above the F_2 peak to the number below n_b.

(b) The ratio n_a/n_b is roughly constant at 3:1 irrespective of time and season during daylight hours. Hence the temperate latitude winter anomaly cannot be resolved by proposing changes of the *shape* of the layer.

(c) At night the ratio n_a/n_b increases as would be expected if the rate of loss of electrons decreased with height.

These results have since been confirmed by several American workers. Bauer and Daniels[5, 6] made moon-echo observations by transmitting the signals from Fort Monmouth, New Jersey and receiving them at the University of Illinois. Because they used only a single frequency f these authors were compelled to make certain assumptions concerning the growth and decay of the ionosphere to resolve the ambiguity in the total number of rotations Ω. Hill and Dyce[42], and Dyce[23] were similarly handicapped, though Millman *et al.*[56] used a high frequency ($\sim$400 Mcps) for which the number of possible values of Ω was limited to two, but then the accuracy of the measurements tends to be poor. Apart from these three American groups, few others have attempted to repeat the early work at Jodrell Bank. The reasons are not very hard to find. In the first place a powerful radar system is required to detect echoes from the moon, and secondly, the advent of the artificial earth satellite has provided (for most organizations) a cheaper means of making these kind of measurements. The difficulties which were anticipated in separating the fading due to the tumbling of the satellite from the Faraday fading were never fully realized, and hence measurements of electron content n_t were made by various groups soon after the launch of Sputnik I.

In the satellite work it is customary to measure the change $\Delta\Omega$ caused by the change in the term $\overline{B \cos\theta \sec x}$ as the satellite crosses the sky. This requires the assumption that the ionosphere be spherically stratified and of equal thickness over the region covered by the motion of the satellite. An additional difficulty is encountered when the satellite moves within the ionosphere, for then the change in n_t caused by its vertical motion can only be separated from that introduced by the change in $\overline{B \cos\theta \sec x}$ if the orbit is known quite accurately. Despite these difficulties, Aitcheson *et al.*[1], Garriott[32], Blackband *et al.*[9], Blackband[10] and others have attempted to measure

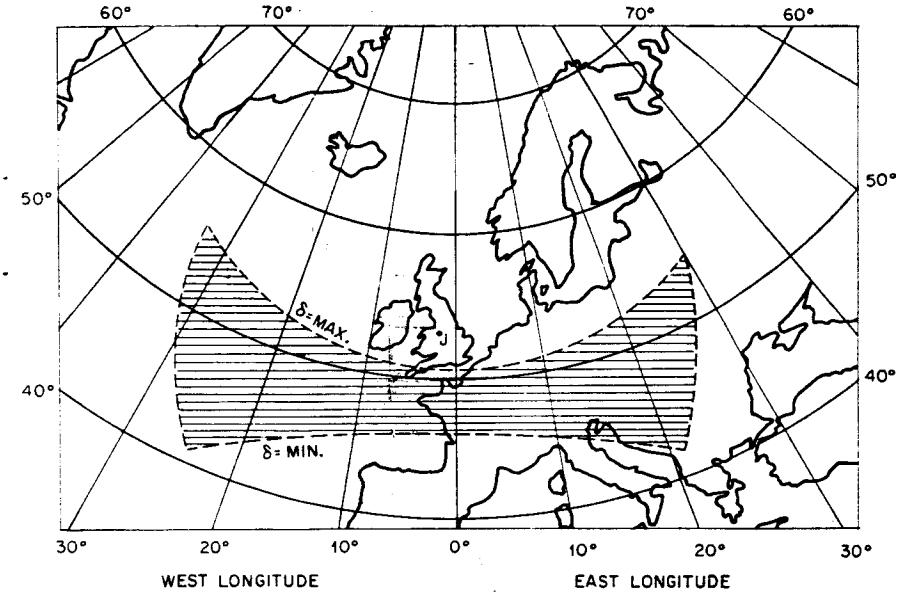
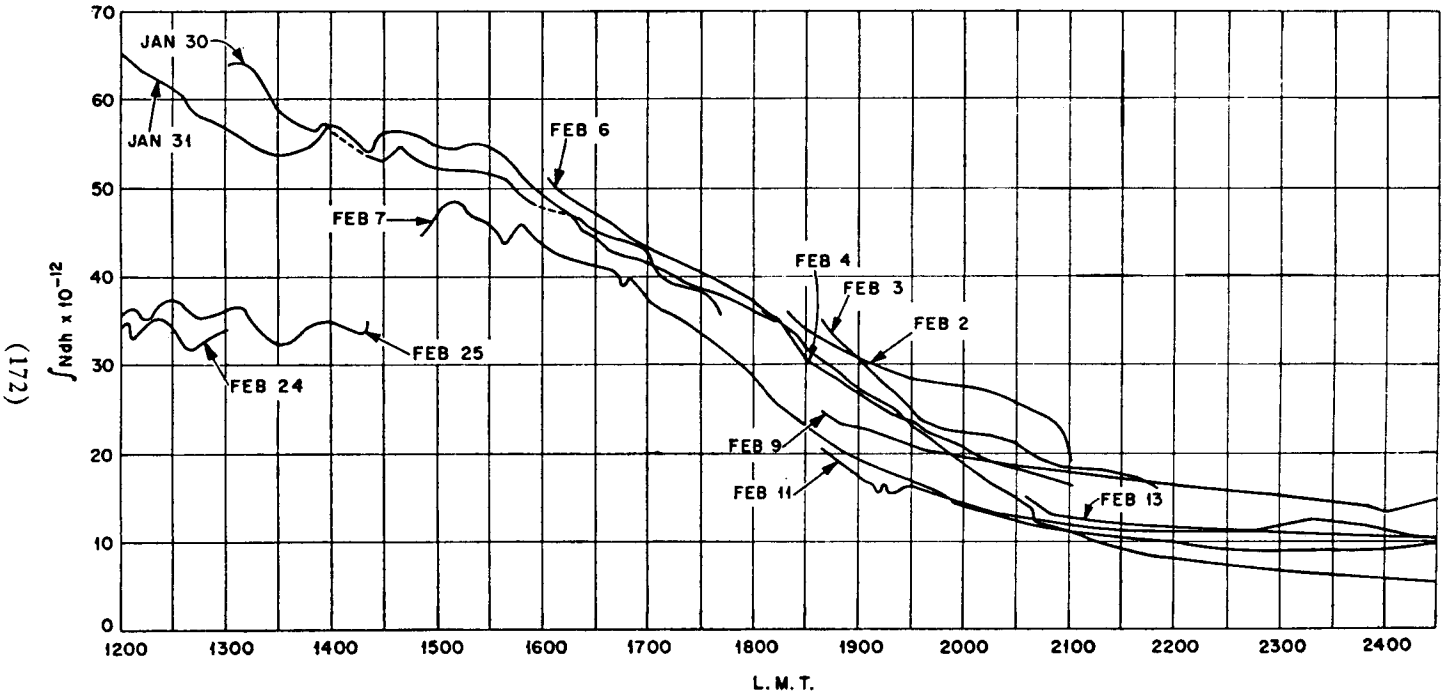


Fig. 1—The shaded region of this map shows the area on the surface of the Earth covered by the motion of the sub-ionospheric point. The variation in latitude is largely caused by the change in the moon's declination over the month.

n_t by satellite observations. Some workers (Hame and Stuart[37, 38], Yeh and Swenson[93]) have made two frequency observations using the 20 and 40 Mcps from 1958 δII to resolve the ambiguity. In general, the results of the satellite observations support those of the earlier moon-echo work, and in addition, have provided some information concerning the distribution of electrons above $h_{max}F_2$ (Garriott[32]).

Much of this work was accomplished using the 20 Mcps and 40 Mcps signals from the Soviet satellites (Sputniks I, II and III). The simplified form of the Appleton-Hartree equation (used to derive (1)) can then no longer be used since these frequencies are not very much greater than the critical frequency of the ionosphere. Yeh[91] and Little and Lawrence[53] have examined the errors which can be introduced by using the simple theory. Yeh finds that at 20 Mcps there may be errors as large as 30%, whilst Little and Lawrence show that the existence of large-scale irregularities, and possibly horizontal density gradients in the ionosphere, introduce large uncertainties. Thompson and Roger[85], who observed the 108 Mcps signals from 1958 α which moves in an orbit outside the ionosphere, also find serious difficulty in making an unambiguous analysis of their records as a consequence of large-scale irregularities. It seems that most of the moon-echo or satellite observations reported so far have been open to criticism for having made certain simplifying assumptions, thereby introducing errors which could not be estimated.



(172)

Fig. 2—The variation of the total electron content of the ionosphere in a 1 cm^2 vertical column.

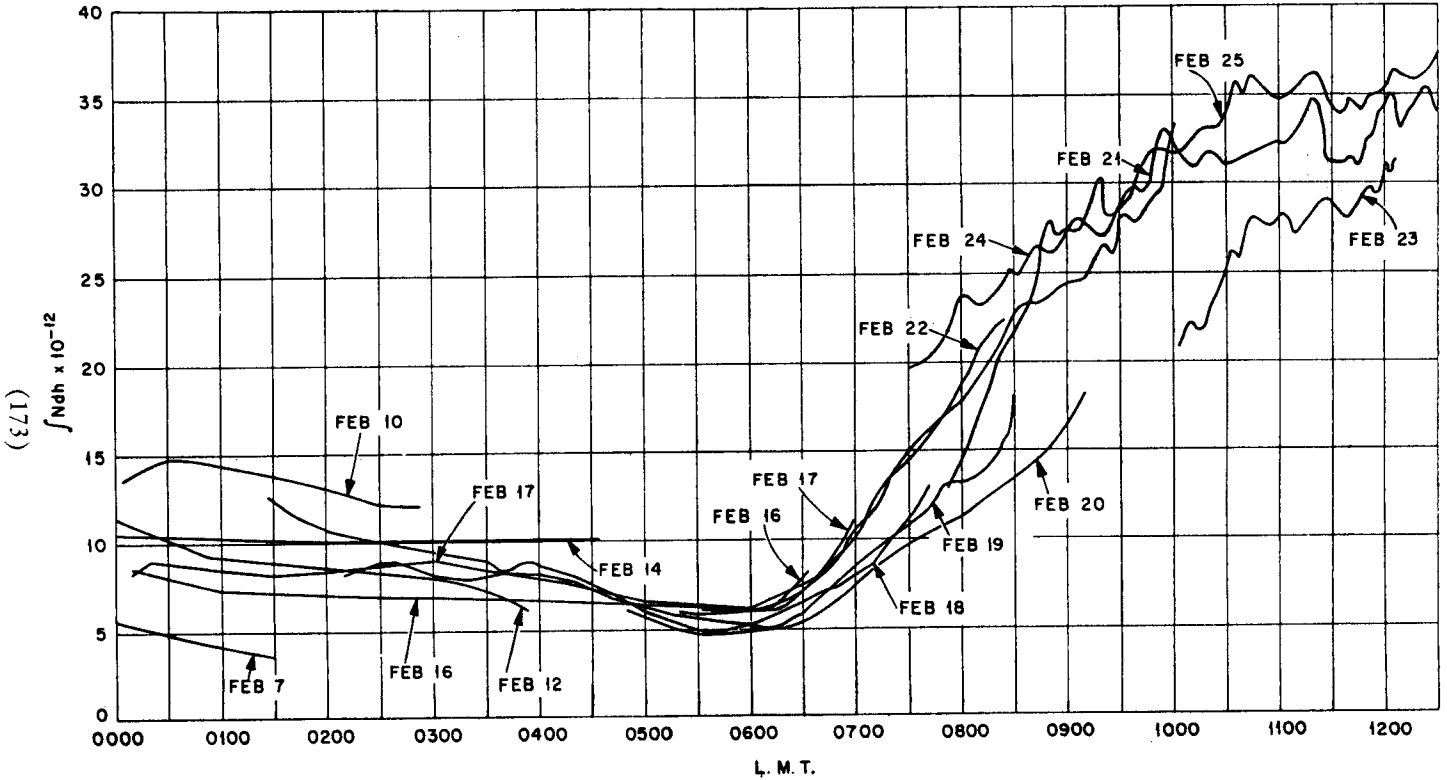


Fig. 3—The variation of the total electron content of the ionosphere in a 1 cm² vertical column.

III. Recent Work at Jodrell Bank

At the beginning of 1960 the workers at Jodrell Bank recommenced their two-frequency moon-echo observations using the 250-foot radio telescope. A full account of some of this work will shortly be published[27] and here I would like to bring out only the more interesting results.

Because these observations are made using two close-spaced frequencies, the ambiguities in Ω are completely removed. Further, the present work makes use of the large radio telescope and this permits continuous observations to be made for as many hours as the moon is above the horizon. The point directly beneath intersection of the ray path through the ionosphere with the 400 km level we shall call the sub-ionospheric point. The motion of moon across the sky will cause the sub-ionospheric point to cover a wide range of latitudes and longitudes. This can be seen in Fig. 1 where the shaded area represents the position covered by the sub-ionospheric point in one lunar month. Because the local time at the sub-ionospheric point will in general be different from that at the observing station, the results, which are shown in Fig. 2 and Fig. 3, have been plotted against local mean time.

From the curves shown in Fig. 2 and 3, it can be seen that to within about 20% the total electron content of the ionosphere repeats itself from day to day. The marked variation in the midday values of n_t observed at the beginning and end of the lunation was largely caused by the seasonal variation, though some of the change may have been due to the variation of the latitude of the sub-ionospheric point.

The irregular nature of some of the curves is quite real, and pronounced random increases and decreases of the number of rotations was found to take place in the daytime. These irregularities have been extracted from the records so that they may be studied separately and some samples are shown in Fig. 4. The peak-to-peak amplitude of these fluctuations was found to be about 3%, and they were observed to occur most readily in the daytime between 09 and 12 L.M.T. (see Fig. 5). The only satisfactory explanation for these temporal irregularities seems to be that they are caused by a wave-like structure in the ionosphere having a horizontal extent of several hundred kilometers which moves with a speed of less than 10 km/second. Such a wave-like motion has been reported by British workers (Bramley and Ross[14], Bramley[15]) and studied in some detail in Australia (Munro and Heisler[60], Munro[61]) by means of reflected signals. The name Travelling Ionospheric Disturbance (T.I.D.) is given to the effect and it has recently been shown to be correlated with Sporadic E.[41] It seems probable that these disturbances are the cause of spatial irregularities that make difficult the reduction of satellite Faraday records[53, 85]. As yet there is no satisfactory explanation for the presence of T.I.D.'s, nor is it known whether they cause the total electron content n_t to fluctuate as huge blobs of ionization move through the beam, or instead if they cause local tilting (1-2°) of the whole ionosphere thereby altering the $\sec x$ term in the product $\overline{B \cos\theta \sec x}$.

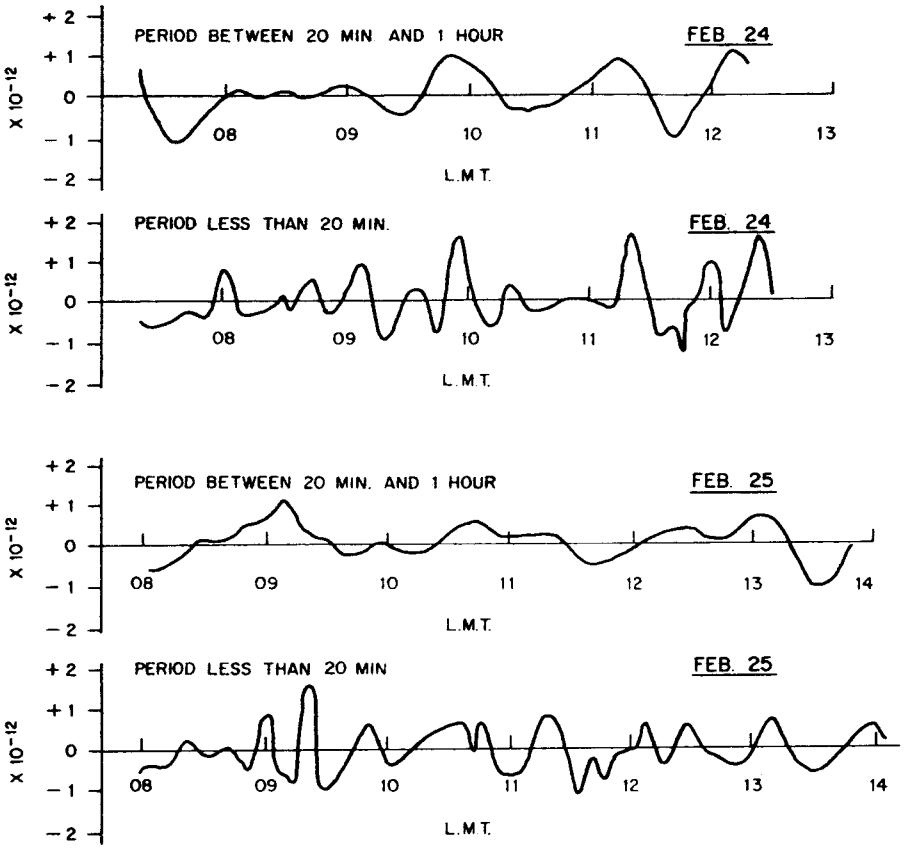


Fig. 4—Irregular variations in the total electron content of the ionosphere.

Half the measurements shown in Figs. 2 and 3 were made when the sub-ionospheric point was over water (see Fig. 1). Thus it is possible to compare only half the results with ionospheric soundings made at the same time. Unfortunately, although there are a substantial number of sounders in Europe, they operate only every hour, and in addition, if restrictions are placed on the appropriateness of sounding results depending upon the distance of the sub-ionospheric point from the nearest sounder, then only a few measurements are available for comparison. However, by analyzing the available critical frequency curves of Slough, Freiburg, Genoa and Graz, values for the number of electrons n_b below the F_2 peak were obtained and the ratio n_a/n_b then plotted in Fig. 6. It can be seen that although there is a considerable scatter in the values observed at night, the ratio was consistently close to 3:1 during the daytime. The mean variation of n_a/n_b was obtained by arbitrarily giving twice as much weight to the points in Fig. 6 which are thought to be accurate and this is shown in Fig. 7. It seems that throughout most of the day $n_a/n_b \leq 3:1$ and during the night the ratio

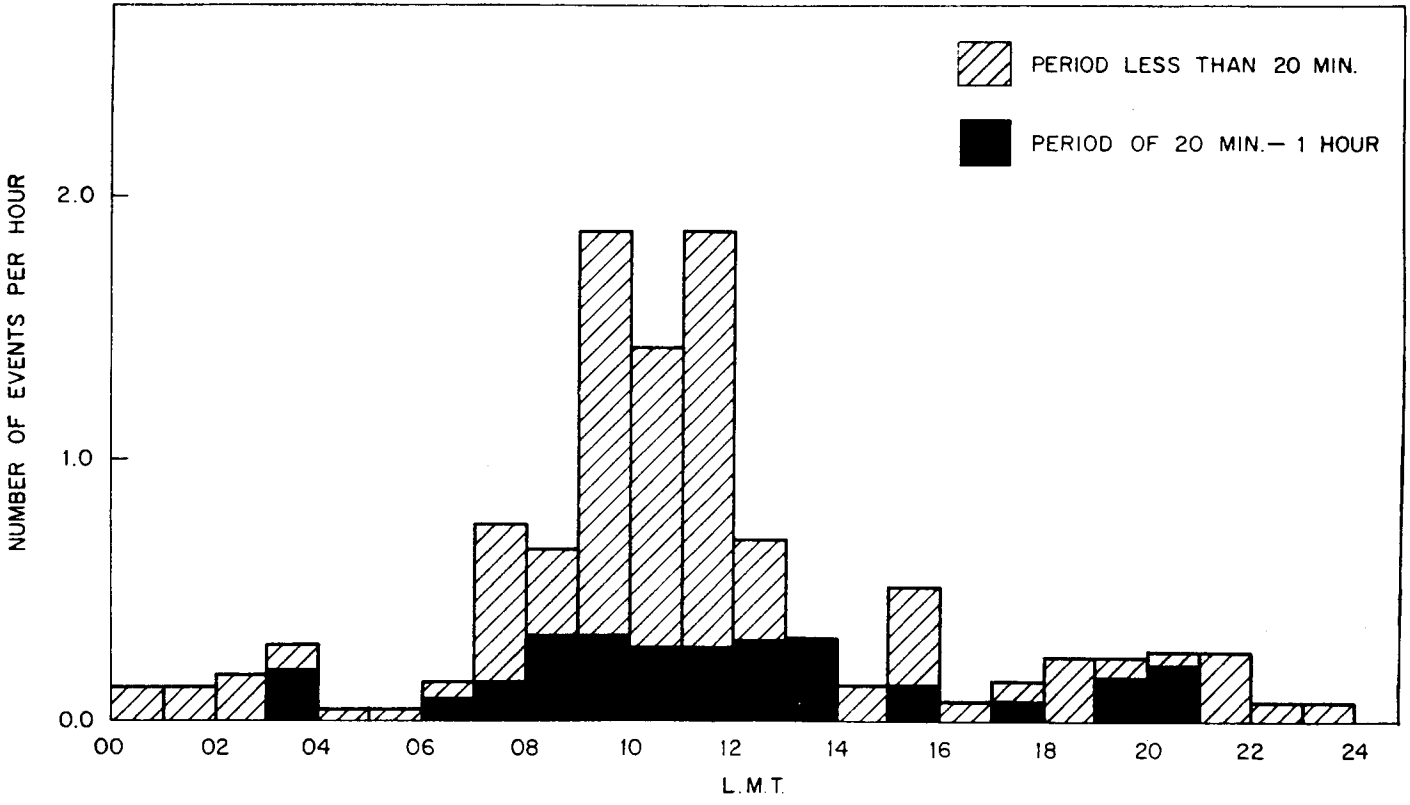


Fig. 5—The frequency of occurrence of irregular changes in the electron content.

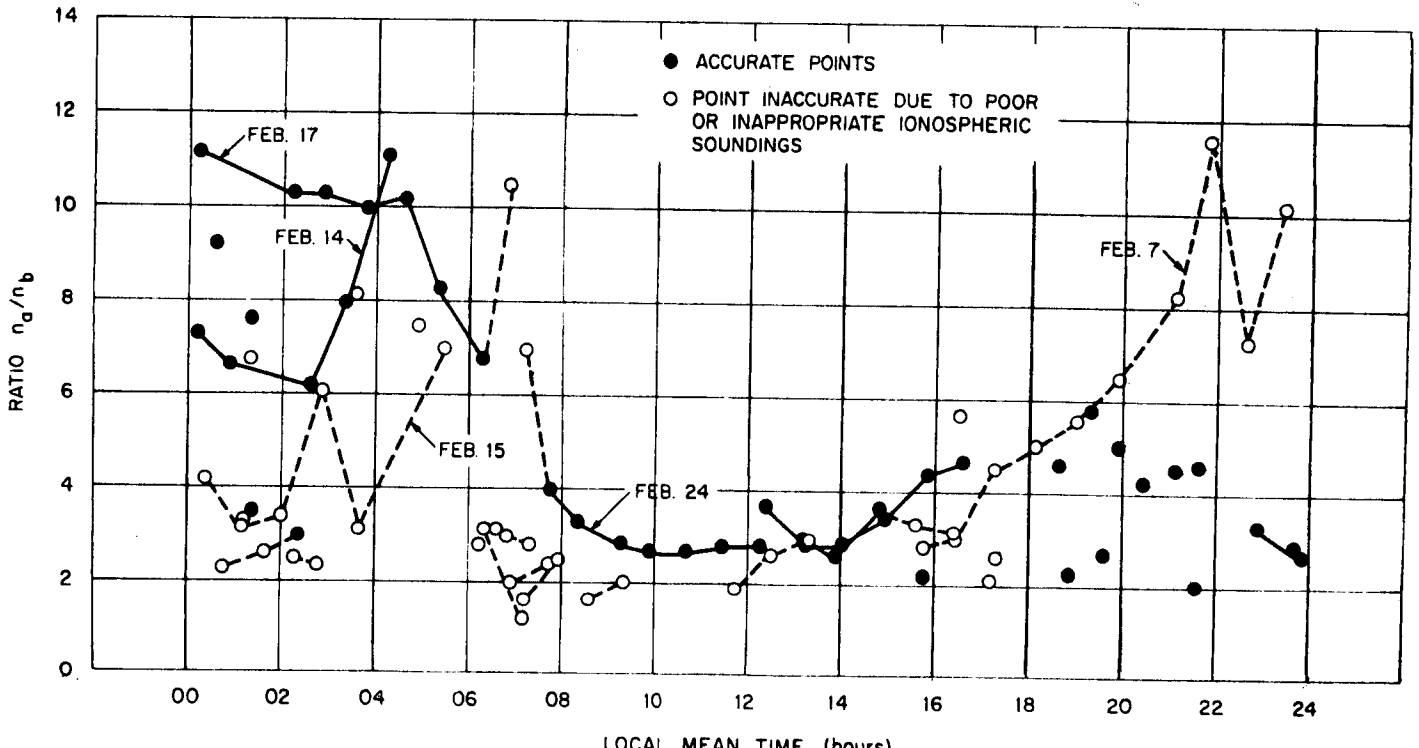


Fig. 6—The values observed in February 1960 for the ratio of the number of electrons n_a above the peak of the E_2 region to the number below n_b.

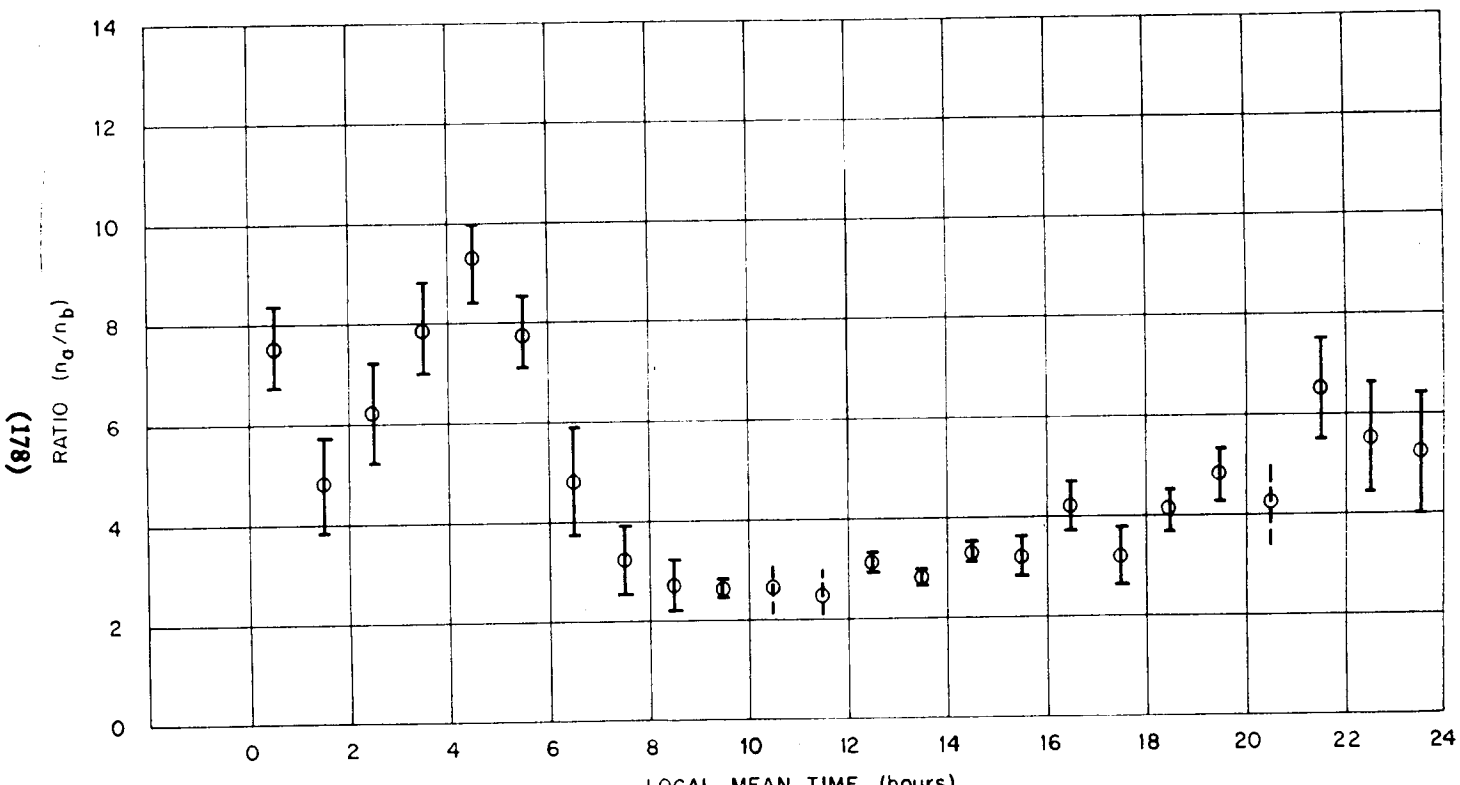


Fig. 7—The mean variation of the ratio n_a/n_b plotted for the points given in Fig. 6 by arbitrarily giving twice as much weight to the accurate points.

increases, reaching a peak just before sunrise in the ionosphere. It then falls rapidly as fresh ionization is produced in the F_1 and lower part of the F_2 layers.

It is possible that some of the scatter in the values of n_a/n_b observed at night was caused by magnetic activity for by chance there was a wide scatter

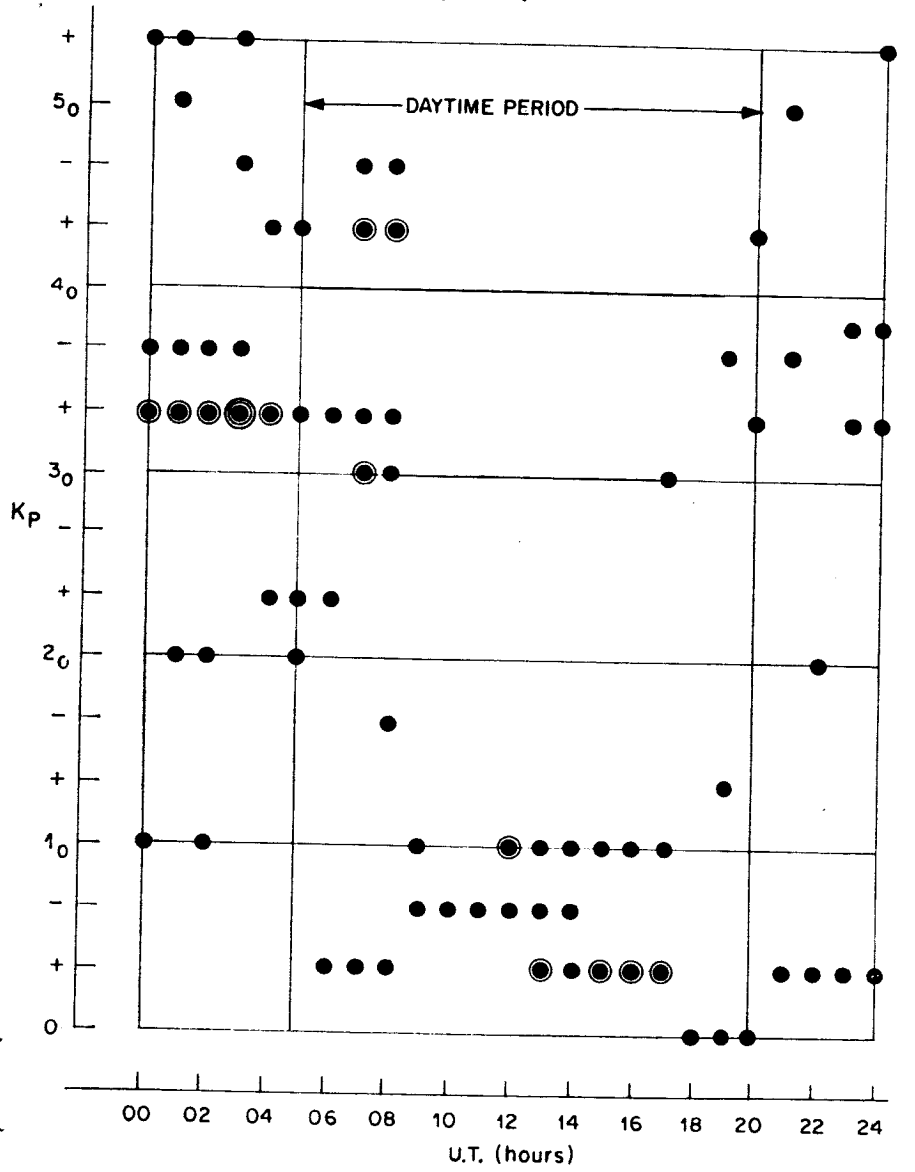


Fig. 8—The values of K_p the planetary magnetic index for all the points shown in Fig. 6. It can be seen that by chance K_p was low throughout the daytime periods of observation.

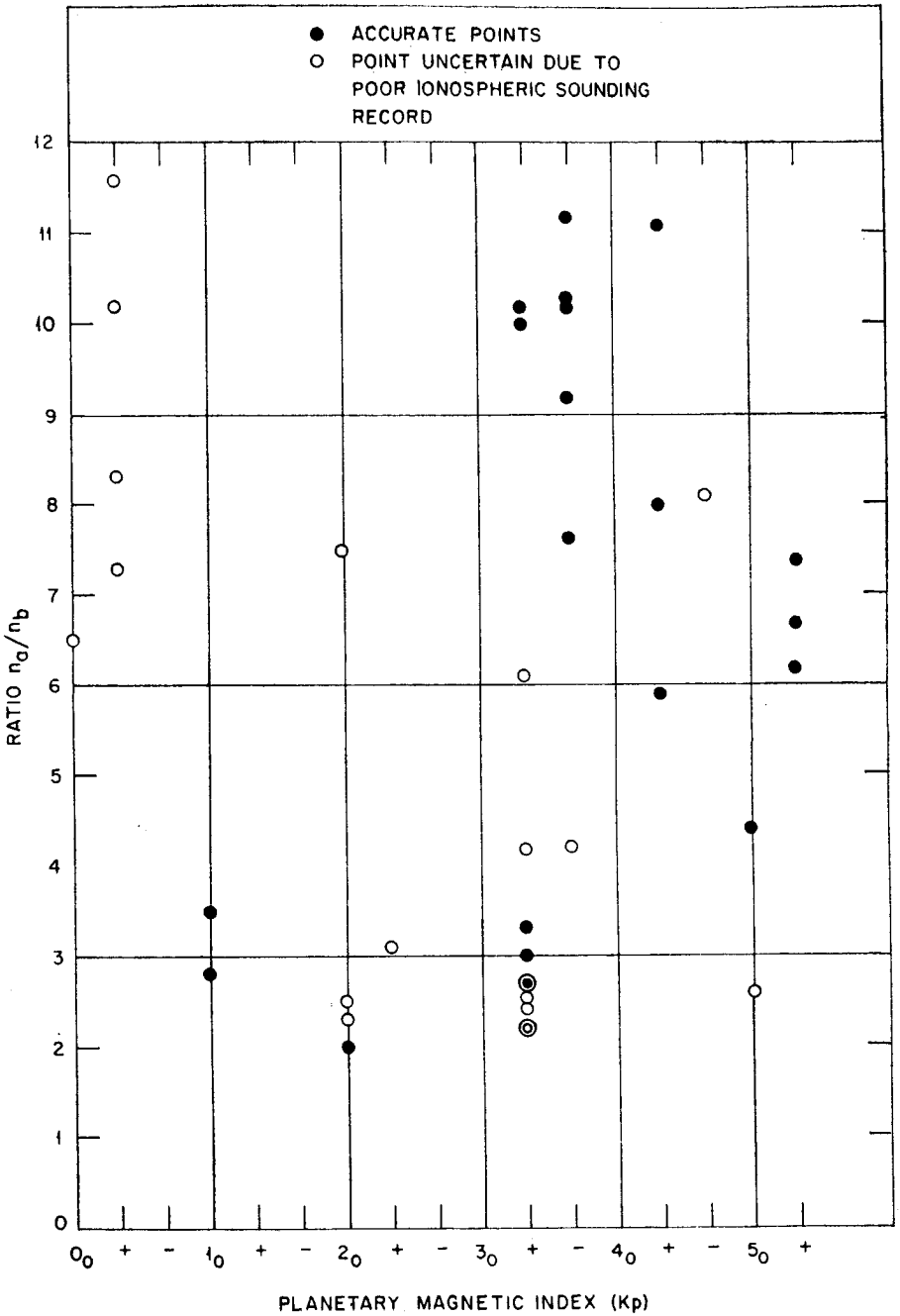


Fig. 9—The values of n_a/n_b (Fig. 6) observed during the night-time plotted as a function of the planetary magnetic index K_p.

in the values of the planetary magnetic index K_p during the night-time observations, but not during the day (see Fig. 8). However, a plot of n_a/n_b against K_p for the night-time hours (Fig. 9) does not indicate any clear correlation between these two quantities.

The ratio n_a/n_b to be expected for a Chapman region would be 2.15. The results indicate a higher ratio than this, though according to Nisbet[66] (also Nisbet and Bowhill[67]) the ratio n_a/n_b obtained in moon-echo work is an overestimate because the values for n_b deduced from ionospheric soundings are systematically underestimated. These authors make this statement on the basis of rocket soundings which showed that the F_2 peak was some 20 to 40 kms above the position predicted by ionosonde data. It is not clear whether this criticism applies to the results reported here, because the reduction methods employed to obtain n_b were those of the ionospheric group at Cambridge, and these are different from those generally employed in America.

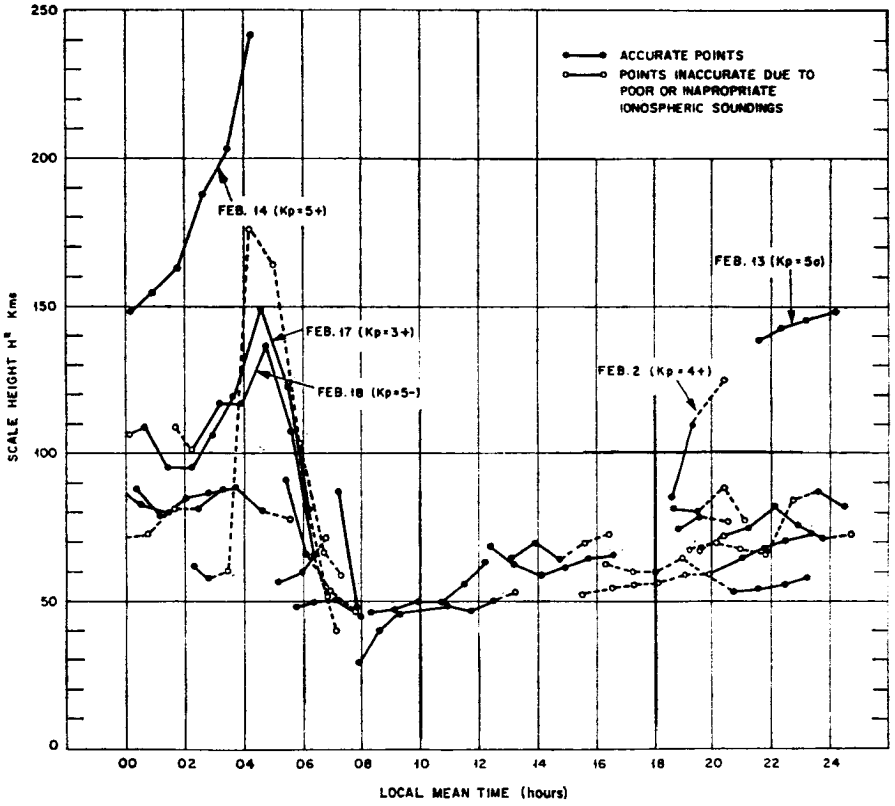


Fig. 10—The scale height H* of a Chapman region which would have the same total number of electrons (n_t) and the same maximum electron density (N_{max}) as observed.

If it were assumed that the electrons in the F_2 region were distributed in a Chapman region, then the scale height H^* of the region could be determined since

$$n_t = 4.133\ N_m H^* \tag{6}$$

where N_m is the maximum electron density. The values of H^* obtained by making this assumption are shown in Fig. 10. If instead one were to compute the equivalent column height H_{eq} given by

$$n_t = N_m\ H_{eq} \tag{7}$$

then values would be obtained which are 4.133 times those of H^* given in Fig. 10. It seems probable in view of the non Chapman-like nature of the F_2 region that the values for the scale height deduced in this way can be considerably in error. Nevertheless, the daytime values observed ($H^* \approx 50$ kms) are surprisingly close to the value for the scale height of the neutral particles H_N at the peak of the F_2 region derived from satellite drag observations[51]. The significance of this result will be considered later.

The night-time values of H^* show a wide scatter and again it seems possible that this might be a consequence of the greater variation of K_p observed for the night-time results. The night-time values of H^* plotted against K_p in Fig. 11 do indicate a positive correlation between these quantities. It seems that on magnetically disturbed nights N_m is depressed, and the values of n_t is not altered. That is, no marked correlation of n_t and K_p can be found (see Fig. 12). The reduction of N_m during times when K_p is high causes the equivalent column height and the value of H^* to increase considerably. We should expect at night that the ionization will be destroyed fastest at heights just below the maximum and thereby cause H^* to increase. However, the increase of H^* with high values of K_p is somewhat unexpected. One plausible explanation seems to be that atmospheric heating (which is observed in the irregular drag on artificial earth satellites) takes place in proportion to K_p*. The actual heating may occur in the daytime (for example, Nicolet[64], Priester *et al.*[72] and others suggest that variation of solar radiation at 304 A^{ϕ} is responsible for these irregular density fluctuations) and then an expanded ionosphere would be formed. Robbins and Thomas[79, 80] have shown that at Slough N_m is reduced on magnetically disturbed days, and because some measure of correlation exists between the day and night values of both N_m and K_p, we might expect that N_m would be low again at night (when the value for K_p would probably still be high) thus explaining these results. However, it is not possible to exclude electrodynamic forces considered by Martyn[54] as a source of this effect.

A summary of the results of the moon-echo observations follows:

(a) The total electron content of the ionosphere shows variations from year to year which can be interpreted as being caused by the sunspot cycle.

*Thus far the drag on artificial satellites has only been shown to correlate with K_p during magnetic storms, i.e., for high values of K_p.

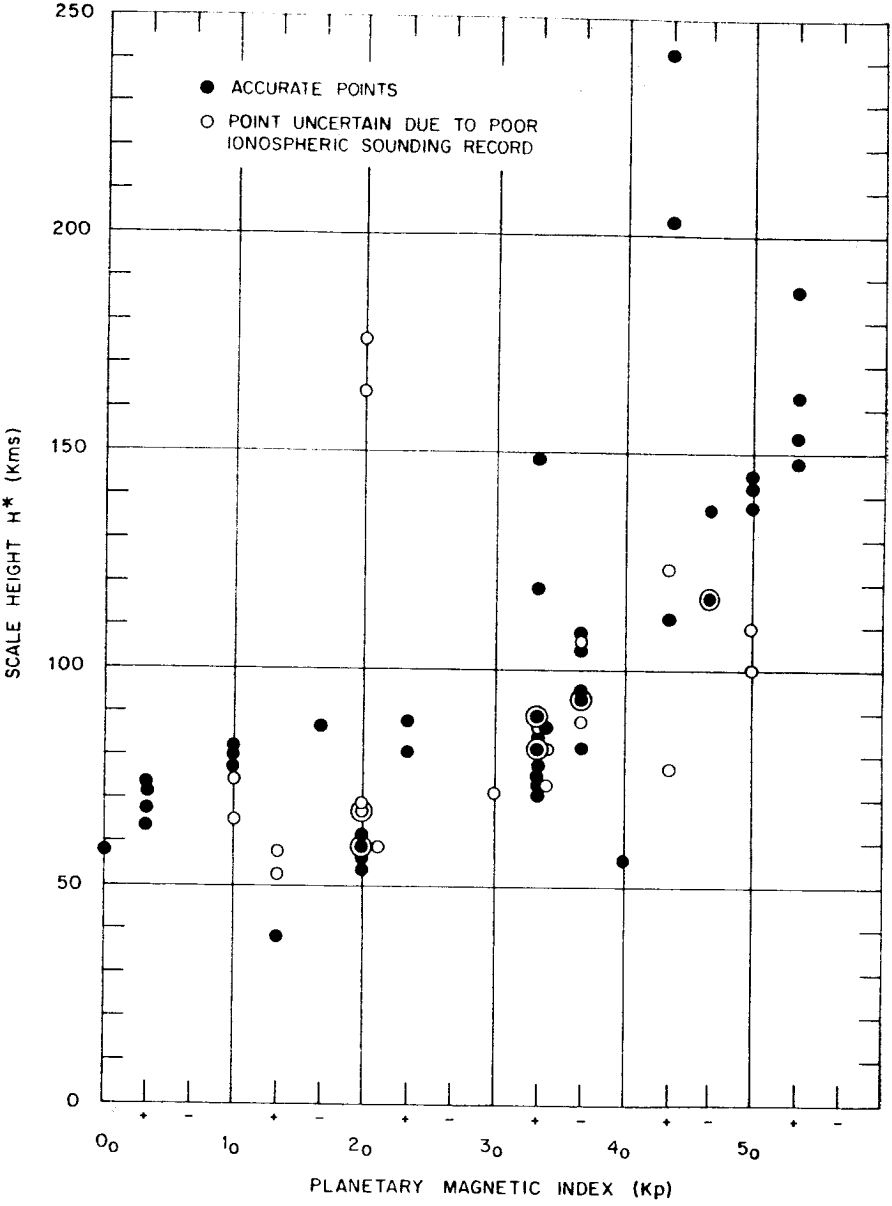


Fig. 11—The scale height H* for the points shown in Fig. 10 plotted as a function of the planetary magnetic index K_p.

(b) The seasonal variation of the total electron content follows much the same pattern as the variation of the maximum electron density.

(c) On a day-to-day basis, the variation of the electron content appears to repeat to within ± 20%.

(d) The electron content is not correlated with small values for K_p though during magnetic storms (K_p high) there is a decrease in n_t[38, 84].

(e) During the daytime (summer and winter) the ratio between the number of electrons above the F_2 peak to the number below is close to 3:1.

(f) At night, in winter, the ratio n_a/n_b shows considerable scatter which may be related to magnetic activity.

(g) At night, in winter, the equivalent column thickness n_t/N_m appears to be well correlated with K_p—at least for values of K_p greater than 3.

(h) During the daytime the scale height inferred from the results is close to that measured for the neutral particles at the maximum of the F_2 by satellite drag observations.

(i) Fluctuations in n_t of the order of ± 3% occur in the daytime in winter and are probably caused by travelling ionospheric disturbances.

IV. Incoherent Backscatter Echoes

In 1958, W. E. Gordon[34] proposed that if a powerful radar beam were directed at the ionosphere a weak but detectable echo from the free electrons there might be obtained. The scattering of X-rays by electrons had long been known in classical physics, but hitherto it had seemed unlikely that practical use could be made of this effect to study the ionosphere. Gordon showed that present day radar equipment is just capable of observing this effect. The intensity of the echo can be computed by considering the electrons to scatter independently with a scattering cross section σ_e given by the square of the classical electron radius,

$$\sigma_e = \left[\frac{e^2}{mc^2}\right] \qquad \text{in Gaussian units.} \qquad (8)$$

i.e., $\sigma_e = [2.8 \times 10^{-13} \text{ cm}]^2$

where e and m are the charge and mass of an electron. The total scattering from any volume containing N electrons would on average be $N\sigma_e$. That is, because the electrons scatter independently the average power returned is the sum of their individual components. An identical result can be reached if, instead of considering the individual electrons, the random irregularities in the electron density caused by their own motion is examined. In this case, the reflection can be considered to arise as a consequence of irregularities in the refractive index of the medium. Again it is found that the echo power should be proportional to the density N[28].

Because the ionosphere would completely fill the radar beam, the echo intensity would be expected to vary with range R only as $1/R^2$ and not $1/R^4$ as for conventional radar targets. Gordon predicted, however, that because of the high temperature of the electrons, there would be a doppler broaden-

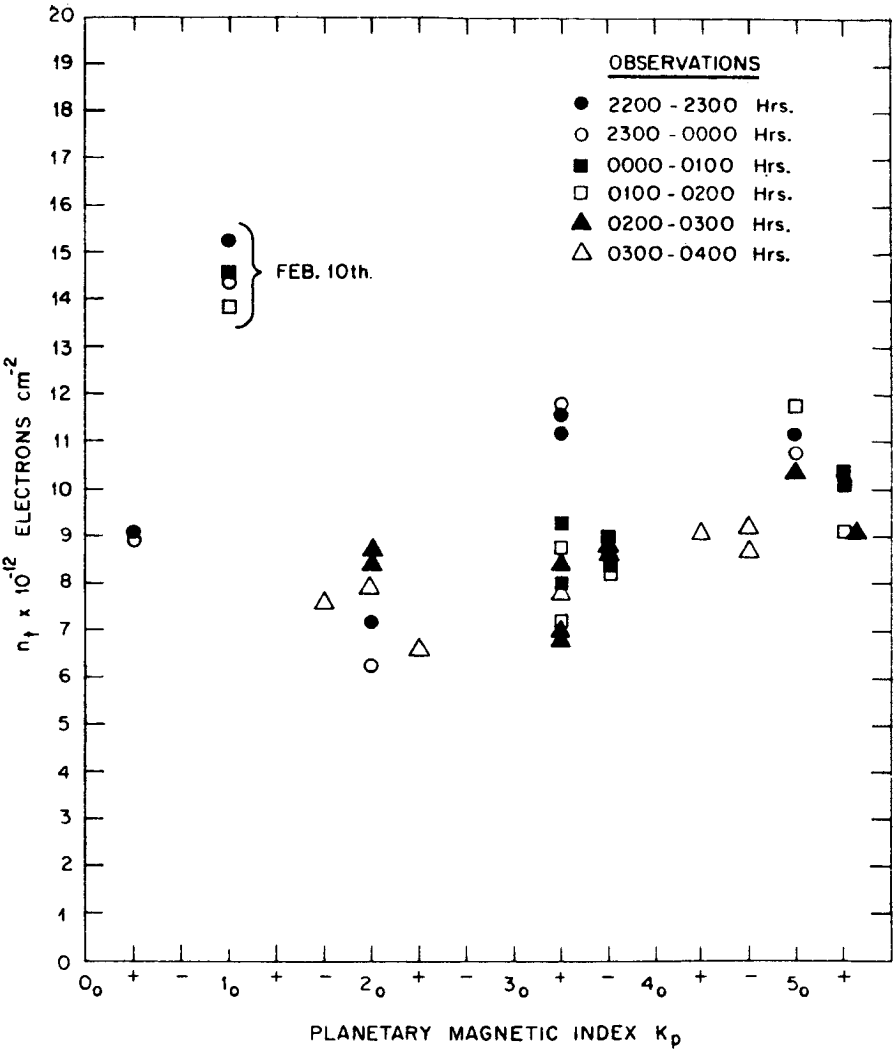


Fig. 12—The values for the total electron content n_t used to derive the values of H* shown in Fig. 10 plotted as a function of the planetary magnetic index K_p.

ing of ~ 100 Kc/s, for a wave at a frequency of 200 Mc/s. Thus the echo power would be distributed over a wide range of frequencies and a broad receiver bandwidth would be required. This would also admit a large amount of unwanted noise into the receiver and cause the signal-to-noise ratio to be poor.

The first experimental observations of these signals came in the same year when Bowles[11] succeeded in detecting the echoes using a radar operating at a wavelength of 41 Mc/s. This equipment used a transmitter power of about

1 MW peak power together with a linear aerial array which had an aperture of 1 acre. Bowles made the important discovery that the doppler broadening of the echo was far less than that predicted by Gordon. He attributed this to the presence of the ions which control the macroscopic random density variations of the electrons in the plasma. This is because the electrons (which have a high mobility compared to the ions) always move in a way which will keep any part of the plasma electrically neutral. The range of these Coulomb forces is characterized by the Debye length λ_D.

$$\lambda_D = \sqrt{\frac{kT}{4\pi\ Ne^2}} \tag{9}$$

where k = Boltzmann's constant; T = the electron temperature; N = the density of the electrons and e the charge on an electron. For most parts of the ionosphere the Debye length is of the order of a few millimetrs, i.e., much shorter than the wavelength of the exploring radio-wave. Hence the Coulomb forces controlling the electrons operate over a shorter range than the wavelength of the radio waves, and as a result the electrons appear to be moving with a velocity close to that of the ions. Bowles[12] modified Gordon's theory to take into account this effect by proposing that the echo could be considered to come from a fictitious particle having the cross-section of an electron but the velocity of an ion. In this case the shape of the doppler spectrum would remain Gaussian in character, but would have a width which is considerably less than that predicted by Gordon.

More recent theoretical investigations of this problem by Fejer[29], Dougherty and Farley[22], Hagfors[36] and Salpeter[81, 82] have shown that this view is not strictly correct. A complete treatment of the problem of the electron-ion interactions yields the result that the strength of the returned signal should be only half that predicted by Gordon, and further that the shape of the spectrum is not Gaussian, but flat topped with "wings" at the edges. The theoretical form of this spectrum is shown in Fig. 13 together with the Gaussian profile for comparison. The "wings" of the spectrum correspond roughly to the doppler shift introduced by electrons moving at the velocity of sound for the ions. Thus the spectrum resembles that which would be caused by highly damped sound waves in the medium. The electrons may be thought of as having two motions. One is their own thermal motion, which being higher than that of the ions would cause a very wide range of doppler broadening. The second motion is that impressed by the more slowly moving ions by means of Coulomb forces. This second motion gives rise to a narrow echo spectrum centered at the radio frequency, and this contains half the total reflected energy.

The importance of the discovery of the narrow form of the central part of the spectrum has been that it becomes possible to make useful measurements of these echoes with an antenna considerably smaller than the one (300 meters in diameter) proposed by Gordon. Thus it has been possible to make measurements using the Millstone Hill radar and these will be discussed later. Both Gordon and Bowles have proposed the construction

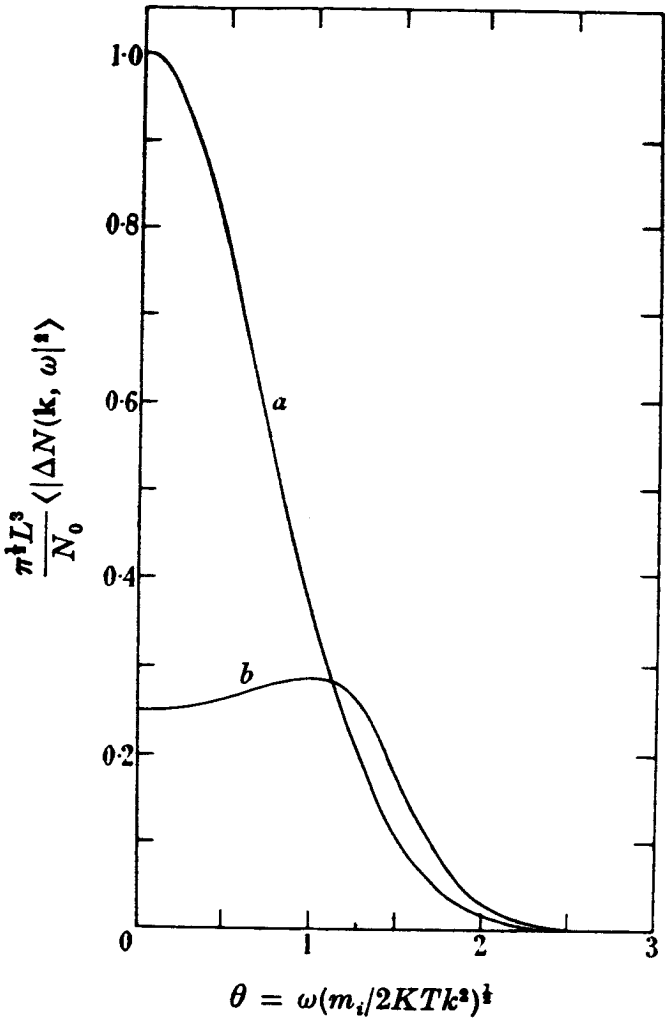


Fig. 13—The normalized power spectrum of thermal density fluctuations (a) for a collisionless gas of neutral particles of mass m_i, and (b) for the electrons in a collisionless plasma in which the ions have mass m_i. Curve (a) has the same shape as the power spectrum suggested by Gordon who supposed that m_i was the mass of an electron thereby giving a spectrum much wider than (b). (After Dougherty and Farley).

of large backscatter radar systems to study these echoes. Gordon's system will be built in Puerto Rico near the town of Arecibo, where advantage can be taken of a large natural depression to construct a 300 meter spherical aperture. This type of surface calls for a complicated feed structure but does permit a limited steering capability. The present design of this system will

permit objects to be tracked to ± 20° from the zenith, with some sacrifice in antenna performance.

The National Bureau of Standards system will be built in Peru jointly with the help of the Instituto Geofisico de Huancayo. This site has been chosen because it lies near the geomagnetic equator. The properties of the ionosphere are distinctly different at the equator because the magnetic field there prevents vertical diffusion of the ions, and as a result the whole ionosphere makes its appearance at much greater heights. Not only will the distribution of electrons in the ionosphere be different at the equator, but the spectrum of the backscattered signal should be different. Bowles has argued that if a radio wavelength is used which is comparable with the radius of gyro-rotation of the ions and if the radio wave is propagated in a direction normal to the field lines, then the backscatter spectrum should show a discrete "line" structure corresponding to modulation at each of the ion gyro-frequencies and their harmonics. Thus a site built on the geomagnetic equator would permit what are essentially mass spectrometer measurements of the ion constituents and their densities throughout the ionosphere. Bowles proposes to build his radar system to operate at a frequency of about 50 Mc/s with a peak power in the range 6-8 MW. The antenna will be a linear array of dipoles covering an area of almost 20 acres. The transmitter may be built in the form of several separate units which will be distributed over this area. Limited steering will be possible by phasing the array, and in this way Bowles hopes to make transit radar observations of the sun.

V. Observations at Millstone Hill

Backscatter echoes can be observed at the Millstone Hill radar facility and these observations have been reported by Pineo *et al.* [68, 69] At best the incoherent echo raises the noise level in the receiver by about 3 db at a delay corresponding to the peak of the F region (Fig. 14). This is observed during the daytime in winter when the critical frequency is high. During the night-time and for much of the summer the echo is considerably weaker. The theoretical strength of the echo P_s can be computed from

$$P_8 \cong \frac{P_t A N \sigma_e (c\tau/2)}{2 R^2 L} \quad \text{watts} \qquad (10)$$

where P_t = peak transmitter power (watts), A = effective antenna aperture (metres^2), N = electron density (per metre^3), σ_e = Thompson scattering coefficient ($\sim 8 \times 10^{-30}$ metres^2), τ = duration of the pulse, L = loss factor of the system, and R is the range of the volume of electrons. The antenna gain does not enter the equation since the total number of electrons in the beam is inversely proportional to the gain and hence this term cancels out.

Bowles' [12] results indicate reasonable agreement between the theoretical and observed echo powers. Pineo *et al.*[70] on the other hand have recently

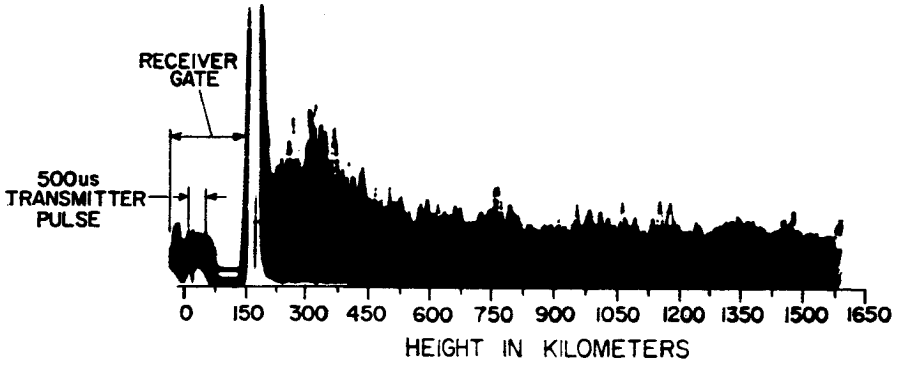


Fig. 14—The appearance of the A-scope display of the Millstone Hill radar when 500 μ. Sec pulses are used to obtain vertical incidence ionospheric backscatter echoes. The receiver has been gated off for a period corresponding to the first 150 Kms height. Between 150 and 200 Kms height there is a strong "ground clutter" echo, and beyond 200 Kms height can be seen the echo from the F region.

reported that at 440 Mc/s the echo power is some 7 db lower than expected. In these measurements N the electron density (at a point near the peak of the F_2 region) was determined using a nearby ionosonde.

Pineo *et al.*[71] have also determined the spectrum of the backscattered power for the region coresponding to the peak of the F_2 region. Their results are plotted in Fig. 15 where they are compared with the theoretical curves of Fejer.[29] Pineo concludes that the best agreement between the observed spectrum is with the theoretical spectrum computed for a ratio of the electron to ion temperature of 2:1. The power under this profile is approximately 2 db less than that to be expected in the case $T_e = T_i$, and hence if the assumption is made that $T_e = 2\ T_i$ the discrepancy between the observed and expected echo power is reduced to 5 db.

The intensity of the echo power as a function of range R yields the distribution of electrons along the path. From such measurements the distribution of electrons with height can be calculated. For the observations made at Millstone Hill the length of the pulse employed in most experiments ($\tau = 250$ microseconds) corresponds to a region 37.5 Kms deep. Hence the distribution of echo power actually observed represents the true distribution of electrons convolved with a slab of this thickness. No attempts have been made to "deconvolve" the results. The echo power is best determined by integrating many samples corresponding to the same range. At Millstone Hill this is accomplished by sampling the output of the receiver by electronic means to provide a series of digital numbers which represent the echo power as a function of range for each sweep of the time-base. These numbers are then summed and stored in a digital computer.

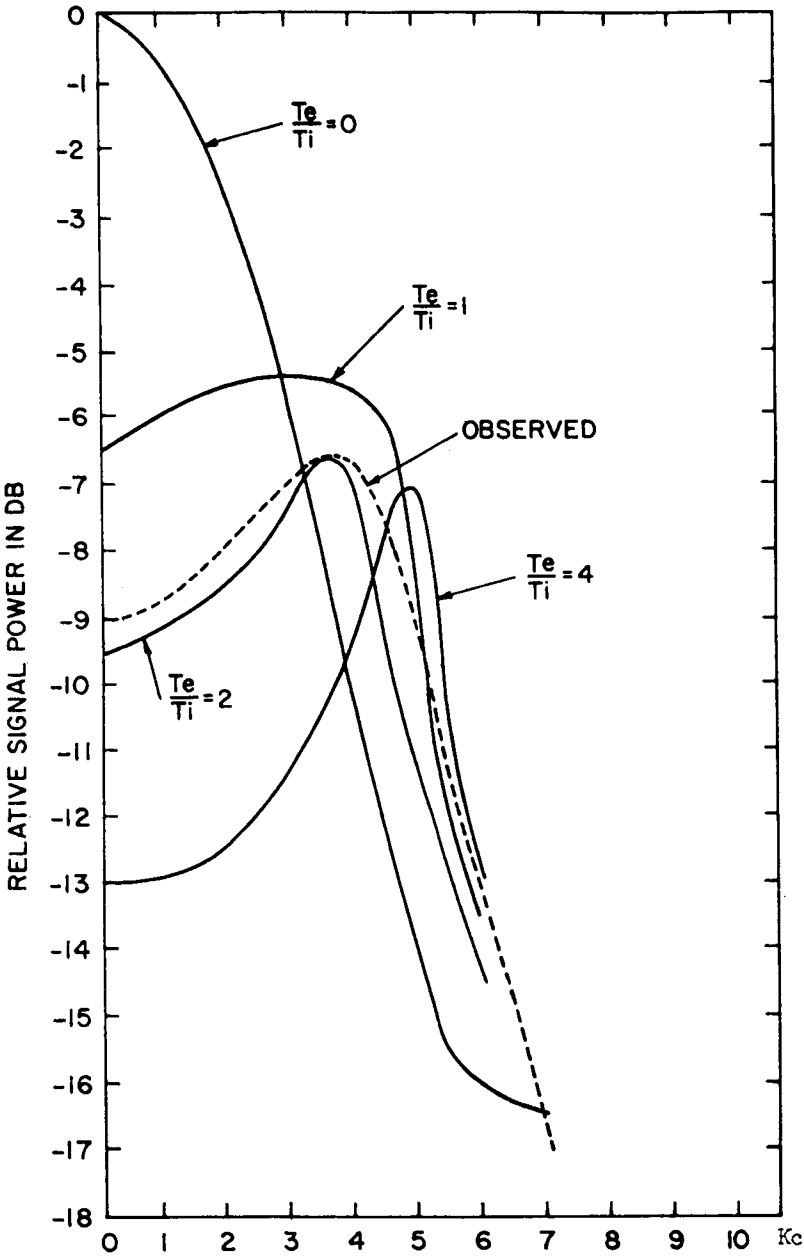


Fig. 15—The theoretical spectrum of Fejer[29] for different ratios of the electron and ion temperatures T_e/T_i are here plotted against a frequency scale of 0-10 KC/s by normalizing their widths so that they are close to the width of the experimentally observed spectrum at 440 MC/s. The spectrum observed by Pineo *et al.*[71] has also been plotted at an arbitrarily chosen reference power.

Figure 16 shows the results of measurements made when the antenna was pointed vertically, and the integration process was continued for 20 minutes. Ground "clutter" echoes prevent measurements at heights below 200 Kms (hence the E region cannot be seen) and beyond 900 Kms the echo is too weak to detect. By lowering the antenna elevation the E region echo can be placed beyond the ground clutter. Figure 17 shows measurements made at an elevation of 15° by integrating for 20 minutes. These two profiles together yield the composite profile for the electron density shown in Fig. 18.

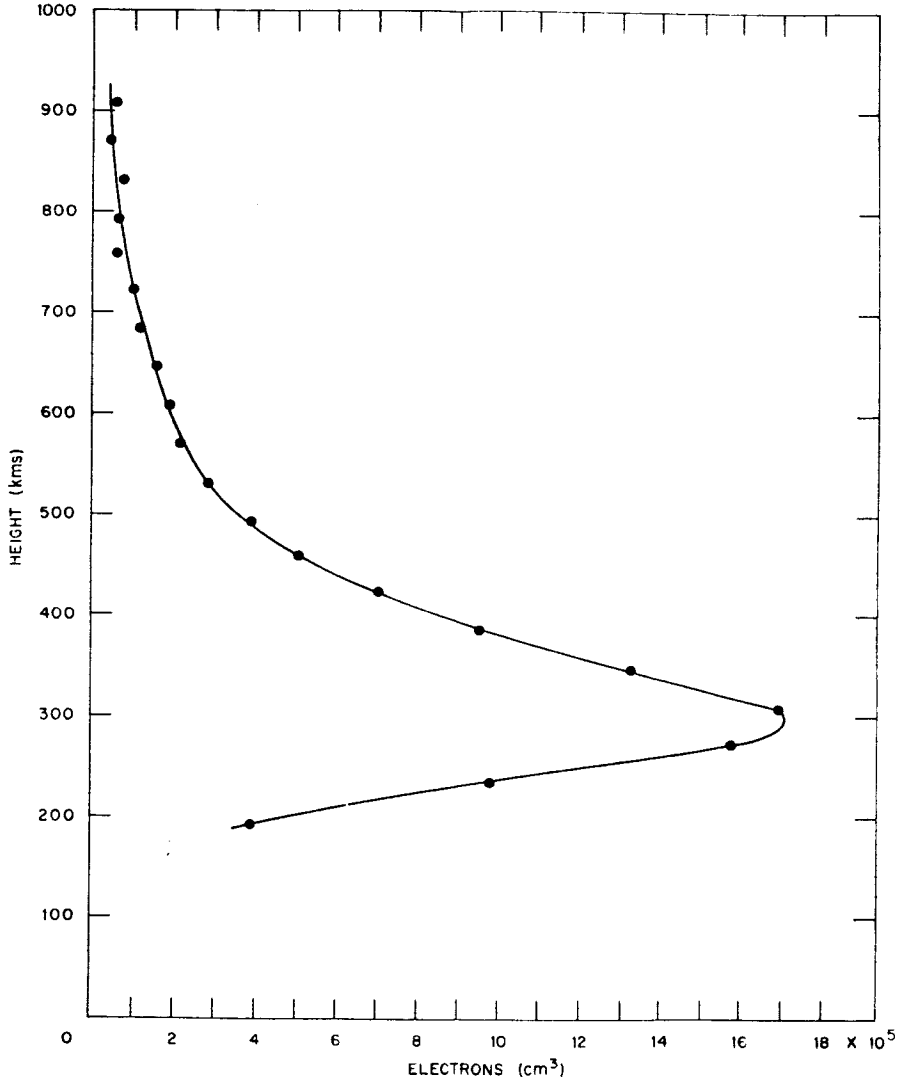


Fig. 16—Backscatter observations on March 18, 1960, between 1400 and 1430 EST.

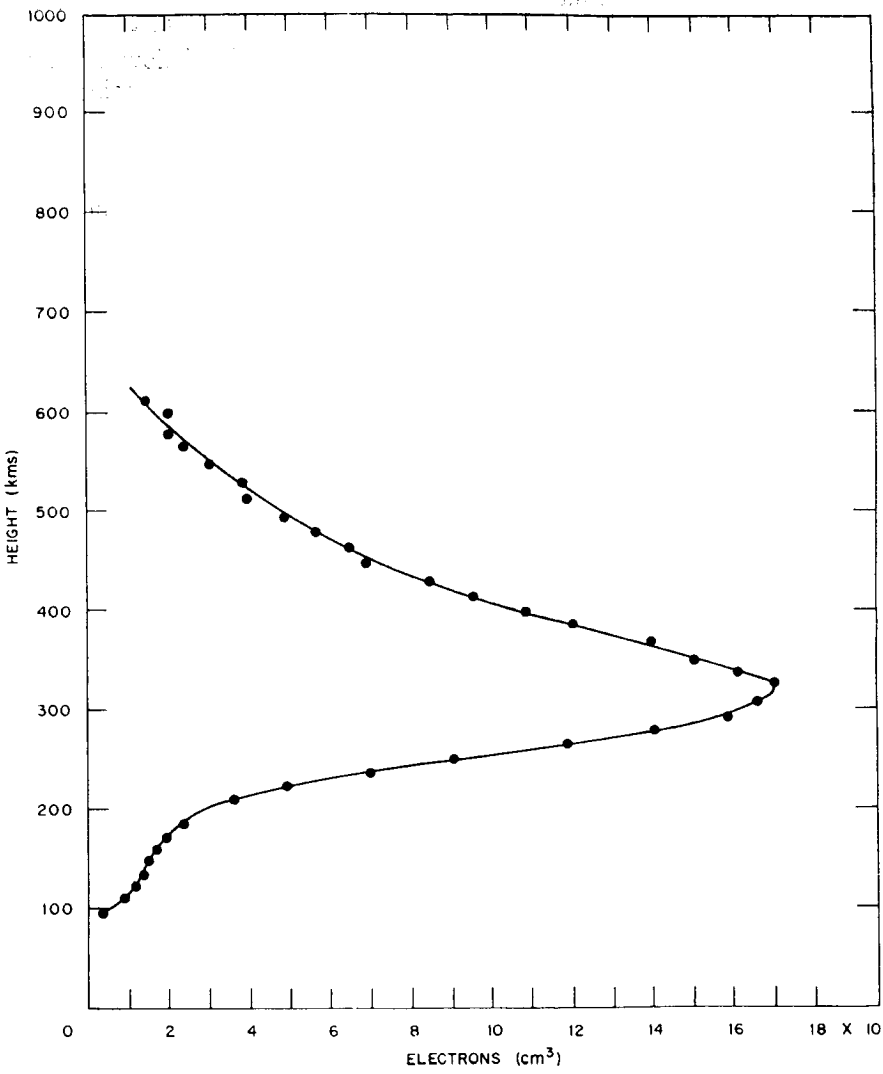


Fig. 17—Backscatter observations on March 18, 1960, between 1520 and 1550 EST.

In view of the discrepancies reported earlier between the observed and predicted echo power, together with the dissimilarity between the echo spectrum and the expected shape (i.e., on theoretical grounds it is to be expected that $T_e \approx T_i$ due to the high conductivity in the region[45]) it seems possible that the electron distributions shown here do not represent the true distribution of electrons. Unfortunately there is little independent information available concerning the complete distribution of electrons in

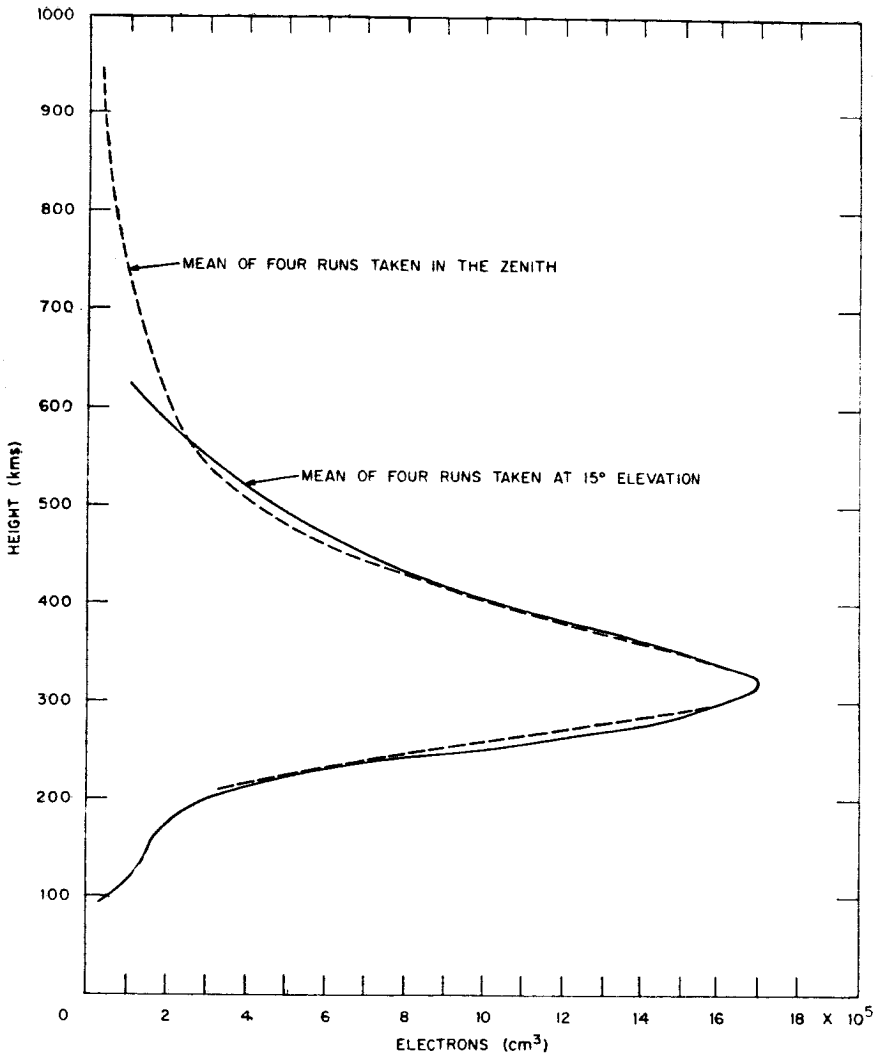


Fig. 18—Backscatter observations, March 18, 1960.

the F region at temperate latitudes. In Fig. 19 the results of two rocket soundings are compared with a backscatter measurement. In Fig. 20 two backscatter measurements are compared with the results of a single rocket profile. Bowles[12] and Millman *et al.*[57] report that there is good agreement between backscatter results and ionosonde density determinations of the lower side of the F_2 region. The evidence at present therefore supports the view that the measurements do represent the true distribution of electrons, at least over the height range which they have been made.

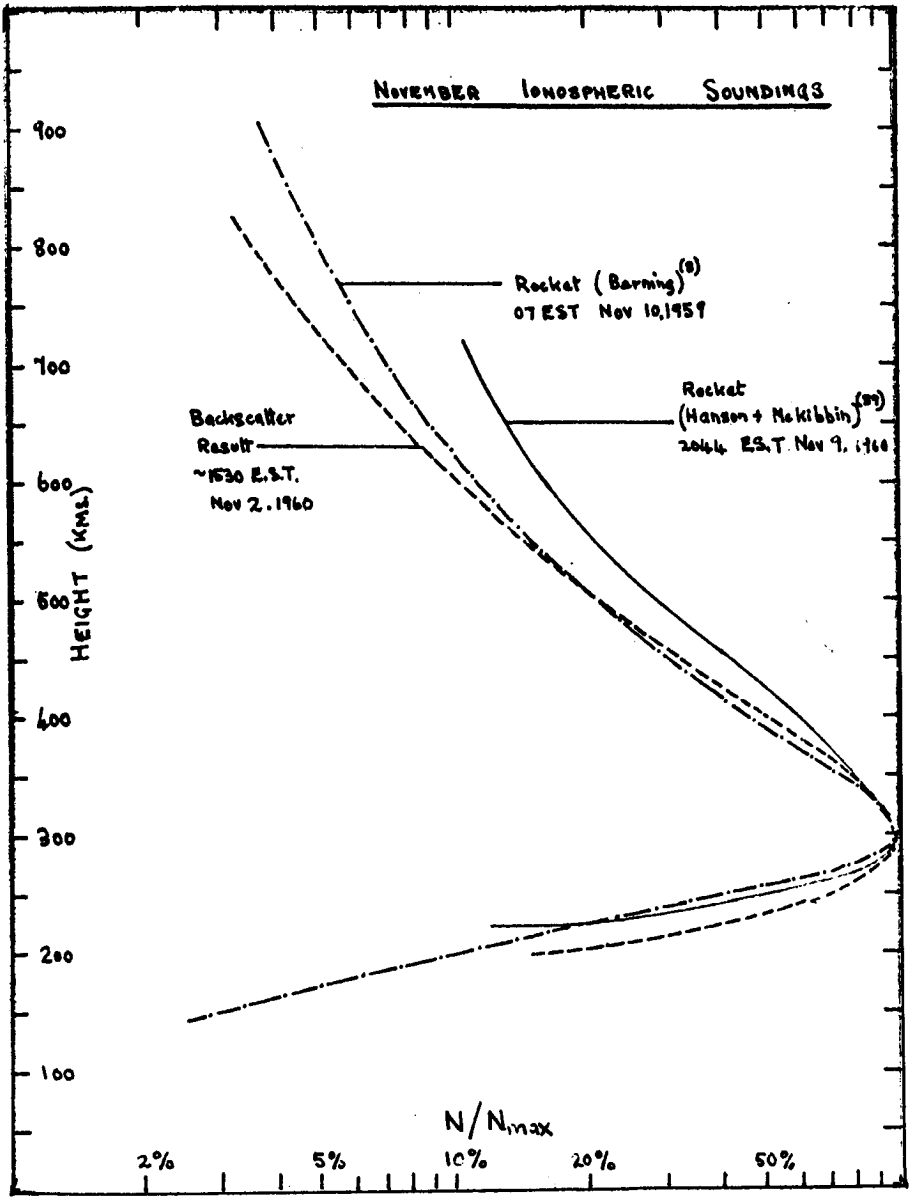


Fig. 19—The results of a single backscatter run during Nov. 1960 compared with the results of rocket sounding from Wallops Island, Va. in Nov. 1959 and Nov. 1960. The heights of the layers have been adjusted so that they all have a maximum at 300 Kms.

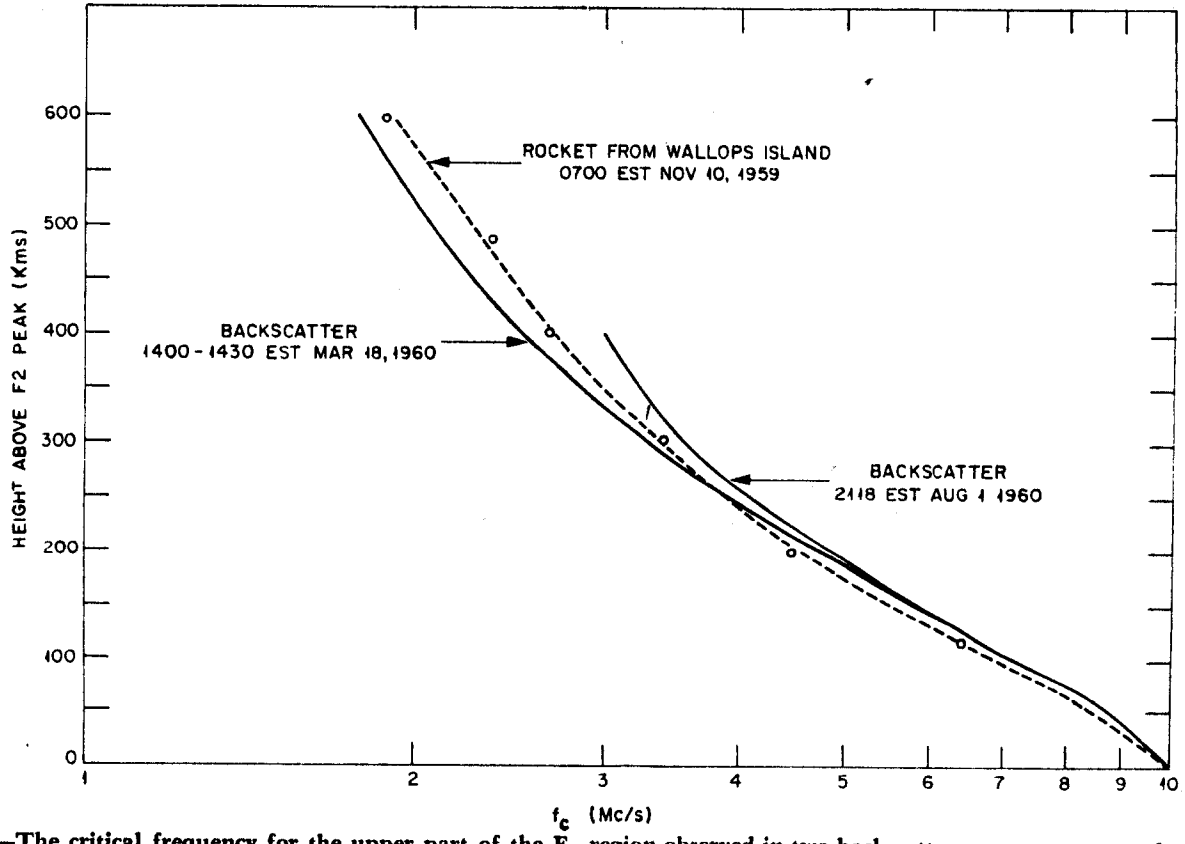


Fig. 20—The critical frequency for the upper part of the F_2 region observed in two backscatter measurements, and a rocket ascent. The maximum critical frequencies have been normalized to 10 MC/s for the purpose of this comparison.

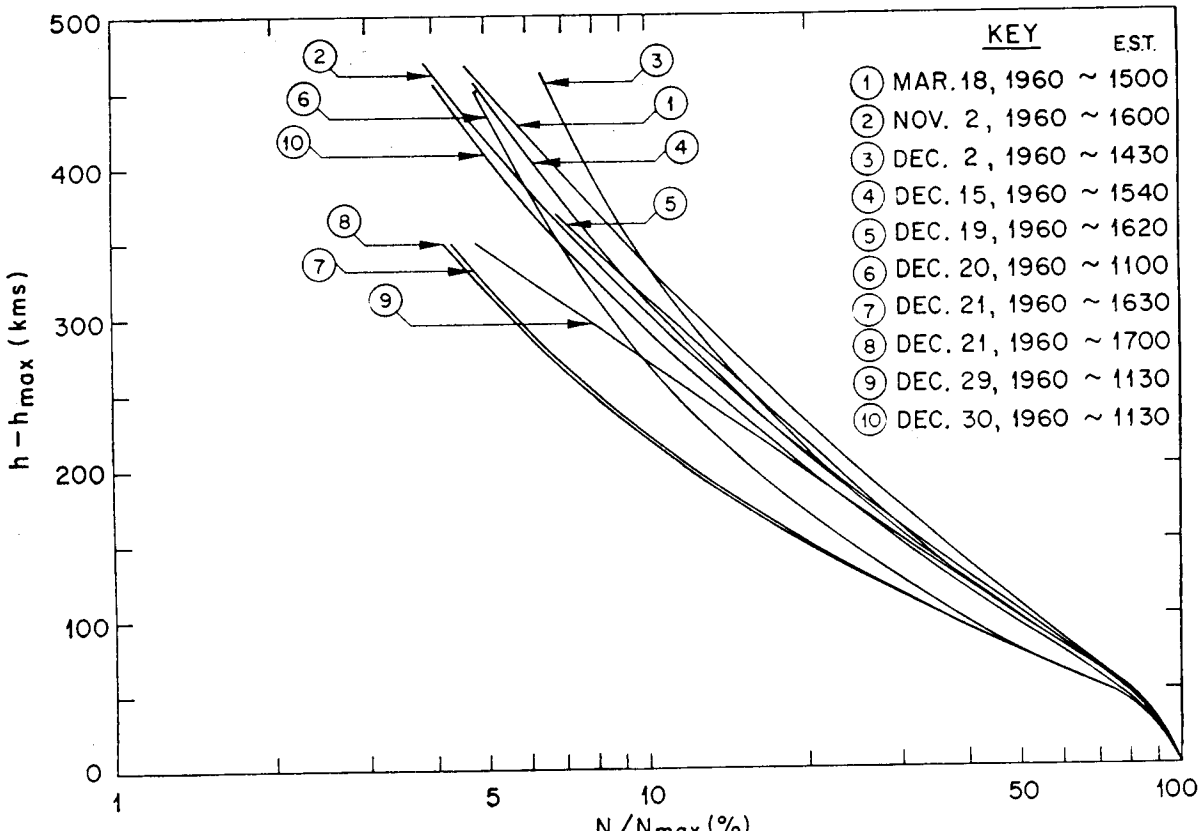


Fig. 21—The electron distribution observed in the upper part of the F_2 layer during ten periods of observation (10-20 minutes integration each) in 1960. All profiles were obtained when the critical frequency lay within ± 1.5 MC/s of 10 MC/s.

Some ten profiles obtained in the manner of Fig. 20 have been selected from the data taken in 1960. The profiles were selected on the basis that the critical frequency observed at Ft. Belvoir at the time of observation was close to 10 Mc/s. The mean critical frequency for the ten runs is actually 10.0 Mc/s and the scatter of their values such that they all lie within $\pm$ 1.5 Mc/s of the mean. In most cases the electron density has only been obtained accurately up to a height of 800 Kms. However, by assuming an exponential decrease of electrons beyond this level with a scale height of 150 kms, the total number of electrons above the maximum n_a was obtained. The number of electrons above 800 Kms, $\int_{800}^{\infty} N\, dh$, estimated by this method was about 0.1 n_a. The ratio n_a/n_b was next computed for each profile and values ranging from 1.3 to 2.3 were obtained. The mean value for all ten profiles was 1.93 with an RMS deviation of 0.26. Much of the scatter in the points may be caused by the method of estimating the exact height of the maximum density which was accomplished by assuming that the distribution near the peak was of parabolic form. The ratio n_a/n_b obtained here is somewhat lower than the value (3:1) inferred from the daytime Faraday results and implies that the criticism made by Nisbet[66] concerning the accuracy of n_b deduced from ionosonde data may well be valid.

The hitherto inaccessible region beyond the peak of the F_2 region is laid open for study by this technique. The density variations for this region obtained from these ten independent periods of observations are shown in Fig. 21. With the exception of two periods on December 21st and one on the preceding day there is considerable similarity between the profiles. In general the density decays to 80% at a height of 50 Kms above the maximum, 50% at 100 Kms above, 33% at 150 Kms above and 22% at a height of 200 Kms above the maximum. This rapid decrease is contrary to most other evidence[76] and suggests that near the peak of the F_2 region the electron density distribution is nearly parabolic[74]. In Fig. 22 the backscatter results (for March 18, 1960) have been added to a curve published by Garriott[32]. In this figure the curve marked *B* represents Garriot's results from Faraday rotation measurements of satellite signals. The other curves represent the results of work done in the U.S.S.R.[2, 35, 47]. It can be seen that the radar results agree best with the absorption measurements of Kazentsev.[47]

The electron density profiles for the region above the F_2 peak shown in Fig. 21 indicate that the electron density decays in an exponential fashion from a height of about 50-80 Kms above the maximum (i.e., beyond 350 Kms height approx.). A similar behavior would be exhibited by a Chapman region. For such a layer the electron density falls off as $\exp\{(\frac{1}{2})[1 - z - (\exp - z) \sec \chi]\}$ where z is the height in units of one scale height. For values of χ less than 60° and $z \geqq 3$ the term $\sec \chi\ (\exp - z)$ can be neglected and then the density falls off with an apparent scale height of twice the scale height of the neutral particles. A Chapman region with a scale height $H = 50$ Kms can be made to fit the present results quite well up to a height of 200-250 Kms above the maximum. This is at variance with a model for this region proposed by Wright[89] who showed that most published observations could be represented by a Chapman region having a scale height of $H = 100$ Kms.

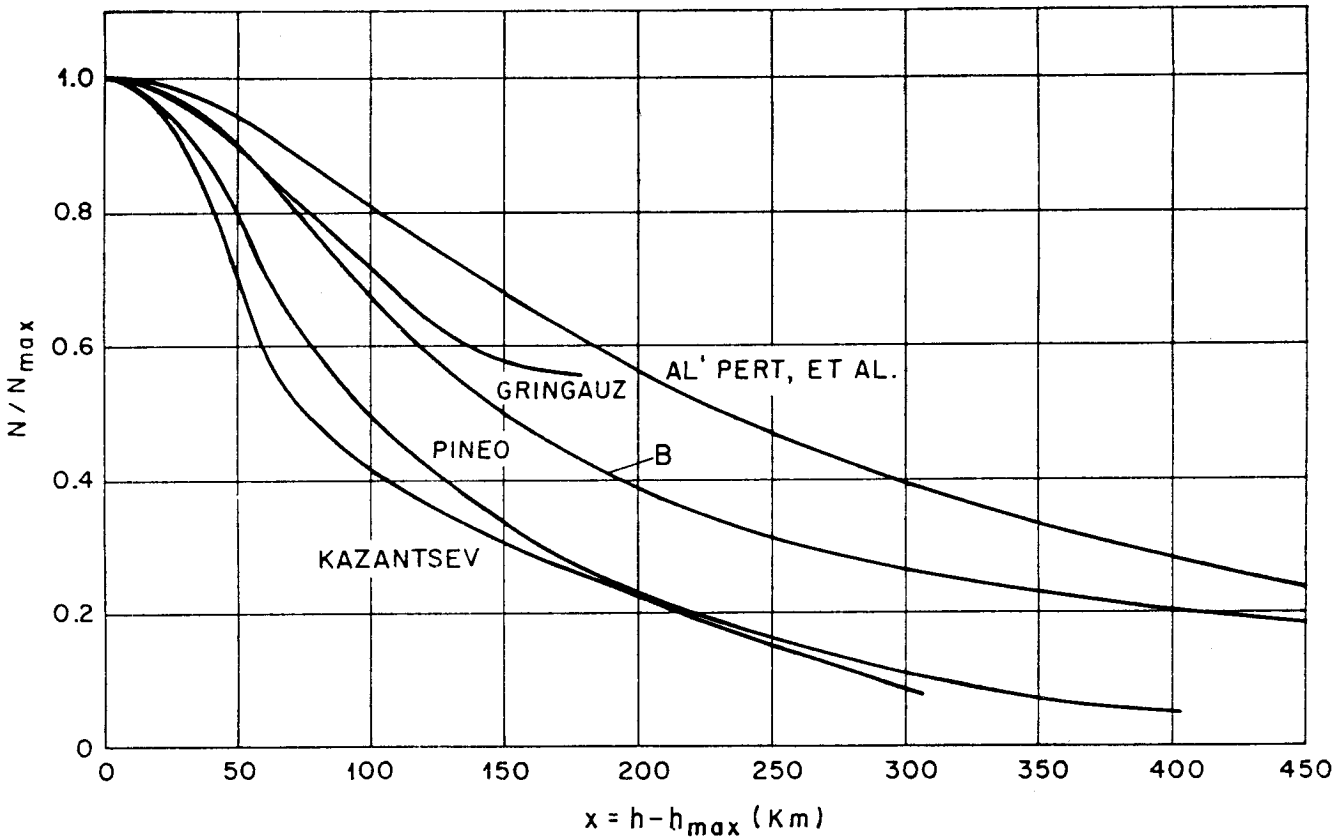


Fig. 22—A comparison of the backscatter profile (marked "Pineo") shown in Figure 18 for the electron density above the peak of the F region, with the satellite results of Garriott[32] (curve B), and the results of measurements made in the U.S.S.R.[2, 35, 47]

The mean values for the scale height at 400, 500, 600 and 700 Kms computed from the curves of Fig. 21 are shown in Fig. 23. If the F_2 region were indeed a Chapman region then the values for scale height of the neutral particles H_N would not vary with height and would be half those shown in Fig. 23. It seems, however, that the F_2 region is not a simple Chapman region and that the electron density above the maximum is governed by diffusion[45, 31, 96]. In this case the scale height of the electrons or the ions is given in[45]

$$H_i = \frac{k\ (T_e + T_i)}{M_i g} \tag{11}$$

where the subscript e denotes electrons and i ions. On the basis of theoretical considerations of the mobility of the particles it has been generally assumed that $T_e = T_i$. If this assumption is made here then values for the scale height of the neutral particles which constitute the parent gas (probably

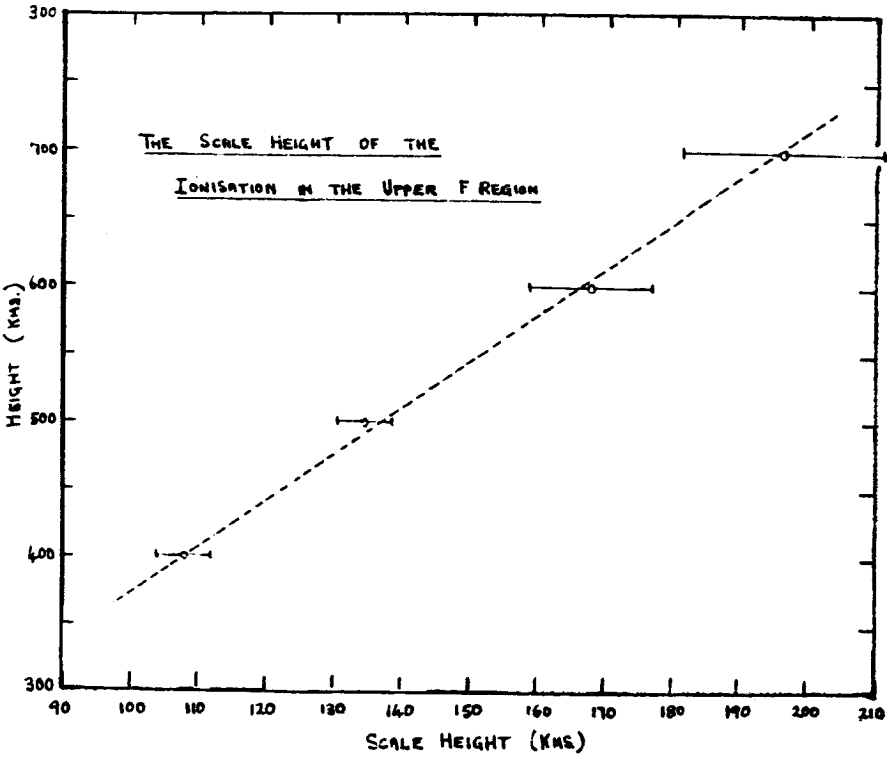


Fig. 23—The mean scale height H_i of the *ionization* in the upper F region obtained from the profiles shown in Fig. 21. The errors given are probable errors. The slope of the fitted straight line indicates that H_i increases at a rate of 0.3.

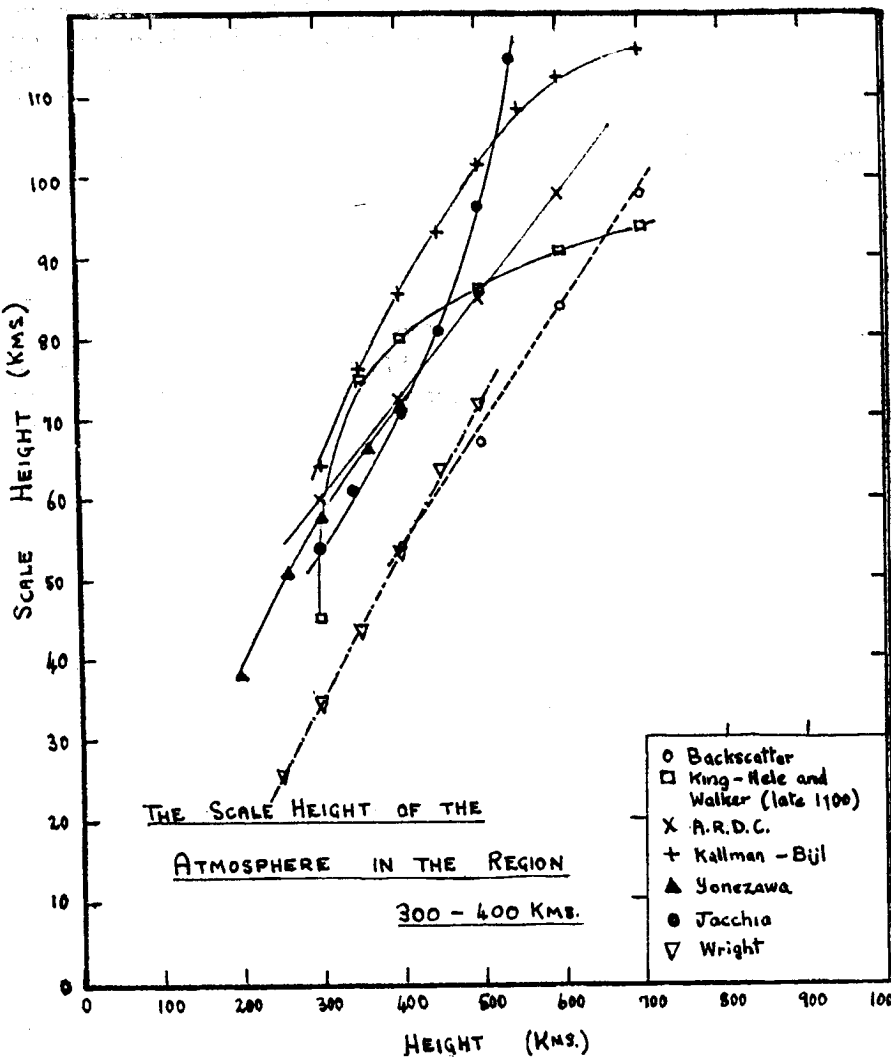


Fig. 24—The mean scale height of H of the *ionizable constituent* in the upper F region obtained from Fig. 23 by assuming that the electron and ion temperatures are equal. Also shown are the ionosonde measurements for the peak of the F region made by Wright.[89] The other curves are satellite drag values for the scale height H_N of all the neutral particles according to various authors.[52, 58, 46, 95, 44]

oxygen) from which the ions are formed can be obtained. These values are plotted in Fig. 24 where they are compared with values obtained from satellite drag observations. In view of the marked year to year variations caused by the sunspot cycle[52] it would seem that a valid comparison of the results can only be made with those obtained by King-Hele and Walker[52].

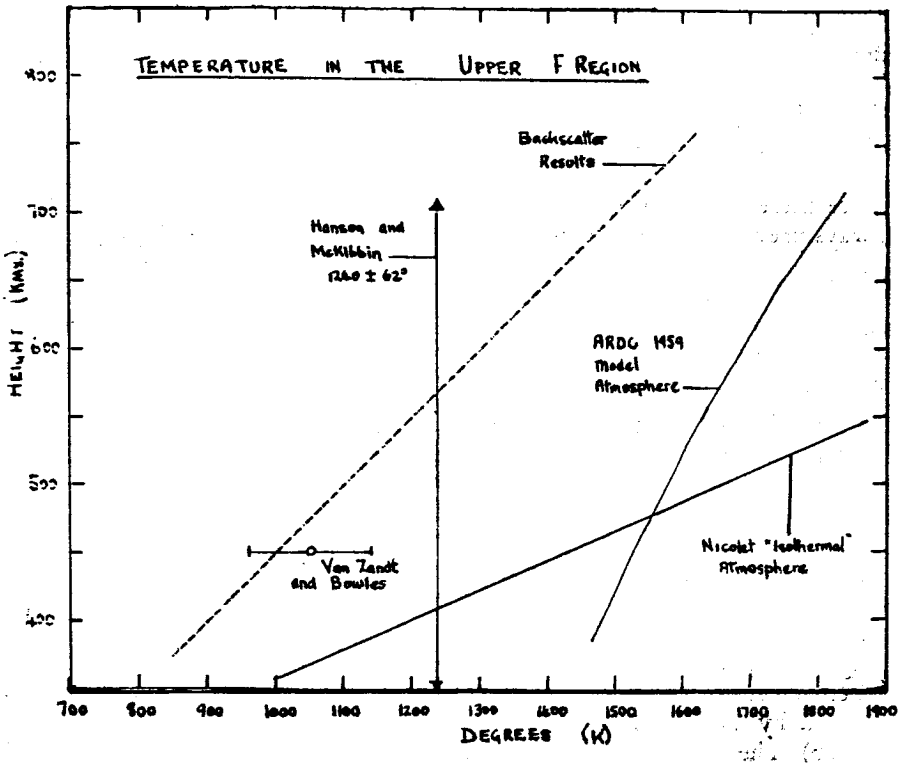


Fig. 25—The temperature in the upper F region obtained from the backscatter results (Fig. 24) by assuming that the ionizable constituent is atomic oxygen. The point marked Van Zandt and Bowles[87] was obtained in the similar way, as was the result of Hanson and McKibbin[39] who used a profile obtained by a rocket ascent instead of backscatter results. The 1959 ARDC model atmosphere[58] temperature is indicated together with a model due to Nicolet[65] which shows the temperature at the point where the particles are distributed such that $\rho(O_2,N_2) = \rho(O)$ and the total density is $\rho = 4.5 \times 10^{-15}$ GM/CM^3 in an isothermal atmosphere. Nicolet[65] makes certain assumptions about conditions at the datum level (150 Kms) and if these are correct it would appear that $\rho(O_2, N_2) = \rho(O)$ at approximately 300 Kms altitude.

These authors consider that their values for the scale height in this region are only accurate to ± 30%. Hence it would appear that though the values for the scale height shown in Fig. 24 are lower than most others there may be no real discrepancy. If the satellite values were found to be lower than the backscatter values this could be attributed to the presence of neutral N_2 or O_2 which would raise the mean molecular mass, and therefore lower the scale height observed in the satellite results. In practice the reverse situation appears to be the case.

(Note added in proof) Recent spectrum measurements of the signals observed at Millstone Hill indicate that only at night is there equilibrium between the

electron and ion temperatures. During the daytime T_e appears to increase above T_i, reaching a maximum near noon when the ratio T_e/T_i is $\sim 1.6/1$. It follows that the backscatter values for the scale height of the neutral particles plotted in Fig. 24 on the assumption that $T_e = T_i$ are over-estimates by up to 20 per cent. This non-equilibrium behavior occurs in a region above 270 Kms at least up to 400 Kms height. It does not occur near 200 Kms altitude, and hence there is a fairly sharp transition between a region where equilibrium always prevails and another region where during the daytime $T_e > T_i$. How far upwards this non-equilibrium region extends is uncertain at the present time as insufficient sensitivity is available to explore in detail the spectra for signals corresponding to regions above 400 Kms altitude using the present Millstone Hill radar equipment.*

We may summarize these results as follows:

(a) The electron density distribution below the peak of the F_2 layer obtained from backscatter measurements agrees well with the distribution inferred from ionospheric soundings.

(b) Above the F_2 peak the density falls off much more rapidly than observed in most other measurements. However the rocket soundings of Berning[8] and Hanson and McKibbin[39] agree well with these results. The less accurate method employed by Kazantsev[47] who made absorption measurements of the signals of Sputnik I also gave results which are in agreement with those reported here. This rapid decrease in density would be predicted on the basis of a more recent model for the ionosphere proposed by Wright[90].

(c) The electron density above the F_2 peak decays in an exponential fashion for all heights above about 350 Kms.

(d) By measuring the gradient of the ion density *vs.* height curve the scale height H_i is obtained. Ten daytime density profiles obtained in late 1960 yield a mean value for H_i of 108 $\pm$ 4 Kms at 400 Kms altitude, and the scale height appears to rise linearly with altitude at a rate of approximately 0.3.

(e) (added in proof) The electron and ion temperatures always appear to be the same near 200 Kms altitude but at 300 Kms and upwards (at least to 400 Kms) the electron temperature exceeds the ion temperature during the daytime.

(f) The electron temperature at 200 Kms height shows little diurnal variation, but at 300 Kms and above there is a large diurnal variation. Also above 300 Kms height the increase in temperature with height is small.*

(g) The electron density distribution near the peak of the F_2 region is close to parabolic in shape. Although the electron density has only been measured up to heights of 800 Kms in most cases, the total electron content can be inferred by assuming that the exponential decrease continues above this altitude. The mean value of the ratio n_a/n_b observed during the daytime in late 1960 was 1.93 $\pm$ .08 ($\pm$ .08 is a probable error). This is considerably lower than the daytime value of 3:1 observed in most Faraday

*These new results are published in Lincoln Laboratory M.I.T. Technical Report 274.

rotation experiments and suggests that n_b computed from ionosonde data is systematically in error.

(h) At the present time there are discrepancies between the predicted and observed values for the total echo power, and also between the predicted and observed spectral shapes. It remains to be seen whether these discrepancies will be confirmed by other workers.

VI. Discussion

At this point it seems pertinent to inquire how these results relate to present day theories for the ionosphere. Because the results principally refer to the F region the discussion shall be limited to a description of the F_1 and F_2 layers. The most widely accepted explanation for the formation of these two layers is that they are caused by the same component of solar radiation. This idea was first proposed by Bradbury[13] in 1938, but only recently seems to have gained widespread acceptance. The difficulty in accepting the Bradbury hypothesis stemmed from the fact that theoretical estimates of the absorption coefficients of the atmospheric constituents and their density placed the maximum rate of electron production at an altitude of 180 Kms. This is approximately the height of the F_1 layer and it was quite difficult to account for the existence of the F_2 layer. In 1947 Bates and Massey[4] suggested that the most important forms of ion production and loss might be

ionization $\quad O + h\nu \rightarrow O^+ + e \qquad (i)$

electron loss
$$\begin{cases} O^+ + O_2 \rightarrow O_2^+ + O & (ii) \\ + & \\ O_2 + e \rightarrow O' + O'' & (iii) \end{cases}$$

The charge-exchange process (ii)(iii) is now thought to proceed at a higher rate than any of the competing processes for the removal of electrons. At one time it was believed that the dissociation of O_2 by means of

$$O_2 + h\nu \rightarrow O' + O'' \qquad (iv)$$

commenced at a height near 90 Kms and was largely complete by a height of 120 Kms[73]. Thus the reactions proposed by Bates and Massey which involve O_2 could hardly take place in the F region. However in 1954 Nicolet and Mange[63] succeeded in showing that although large scale dissociation of O_2 must commence at about 100 Kms altitude, O_2 would still be found at heights consideraly higher. This is because molecules will diffuse upwards from the lower regions to attempt to reform a distribution of molecules in hydrostatic equilibrium (i.e., having an exponentially decaying density). Thus the actual number of O_2 molecules in the F region would depend on the relative rates of diffusion and dissociation. Where the density is high {i.e., $N(O_2) \geqslant N(O^+)$} the rates of reactions (ii) and (iii) will be controlled by the number density N_e of the ions and electrons. Thus

$$\frac{-dN_e}{dt} \propto N_e^2 \qquad (12)$$

and hence the loss rate depends on the square of the electron density, as it would if simple recombination were responsible.

$$O^+ + e \rightarrow O + h\nu \qquad (v)$$

Where, however, the density of the oxygen molecules is less than that of the ions we should expect the rate of reaction (iii) to control the loss of electrons and the loss rate would then become

$$\frac{-dN_e}{dt} = \beta N_e \qquad (13)$$

where β is an attachment coefficient, which decreases with height. In this case the rate of loss of electrons is proportional only to the density and not the square of the density.

By means of analysing ionosonde records for the period following sunset Ratcliffe *et al.*[75] were able to show that in the F_1 layer $-dN_e/dt \propto N_e^2$ and in the F_2 layer $-dN_e/dt \propto N_e$. This piece of research provided strong evidence for the theory outlined above, and suggested that the transition from the region where oxygen molecules are abundant to where they become scarce must lie at about 200 Kms. It should be noted, however, that an alternative charge exchange reaction could produce the same results, namely:

$$O^+ + N_2 \rightarrow NO^+ + N \qquad (vi)$$
$$NO^+ + e \rightarrow N' + O' \qquad (vii)$$

It is now supposed that the peak ion production occurs at about 200 Kms (in the F_1 layer). The F_2 layer is then a subsidiary layer brought about by the fact that the electrons in the upper part of the F_1 have a lifetime considerably longer than those lower down. It is evident that if the density of the neutral particles which are responsible for the attachment process decreases with altitude in an exponential manner, then the lifetime of the electrons will *increase* with altitude in an exponential fashion. On the other hand the production of ionization will decrease exponentially with altitude, and hence when equilibrium conditions have been achieved we might expect the electron density above the F_1 layer to increase to some value and then remain constant with height. Yonezawa[94] and others have shown that this does not occur because there is downward diffusion of ions which try to set up an equilibrium distribution of their own. Because the Coulomb forces between the ions and electrons act to keep the gas electrically neutral the electrons are forced to follow the motion of the ions, and the resulting distribution has an exponential decay with a scale height given earlier in Eq. 11.

$$H_i = \frac{k(T_e + T_i)}{M_i g} \qquad (11)$$

About one scale height below the peak of the layer diffusion is thought to be negligible while at all altitudes beyond one scale height above the maximum diffusion is overwhelmingly important. Early estimates[43, 30, 55] of the rate of

diffusion were too high and indicated that it would be difficult to maintain a bank of ionization at levels of 300-400 Kms. However Dalgarno[20] has reported from laboratory experiments that an ion (O^+) will diffuse more slowly through its parent gas (O) than through an entirely different gas (say N). Thus Ferraro[31] concludes that there is now no conflict between the theoretical rates of diffusion and the rates necessary to permit the existence of the F_2 region. Thus it is clear the backscatter results, which show that a diffusion equilibrium distribution is achieved at a height of one scale height above the maximum, lend considerable support to the theory outlined above.

This theory also succeeds in explaining the geomagnetic anomaly since on the magnetic equator diffusion is inhibited.[94] In consequence the ionosphere there makes its appearance at greater heights and ionization can diffuse only sideways along the lines of force and appear at geomagnetic latitudes of $\pm$ 10° where anomalously high values of the critical frequency are observed.

Yonezawa's[96] treatment assumes that the neutral constituents of the atmosphere are in diffusive equilibrium above some datum level (taken to be 180 kms) and that the scale height H of the ionizable constituent (taken to be O) increases with height linearly at a rate Γ. That is

$$H = H_0 + \Gamma(h - h_0) \qquad (14)$$

Where H_0 = scale height at the altitude of the datum level h_0 (= 180 kms). Then since the presence of O_2 determines the loss rate B this is given in

$$B = B_0(H_0/H)^{1+2/\Gamma} \qquad (15)$$

Excellent agreement is found between the shape and height of the layer computed by Yonezawa for $B_0 = 5 \times 10^{-3}$ per sec., $\Gamma = \frac{1}{8}$ and the mean backscatter profile. Unfortunately Yonezawa used rates of diffusion for the ions which are now thought to be a factor of 4 too high.[31] Thus a detailed comparison of the theory and experiment at this stage is hardly justified. Yonezawa has, however, shown that the influence of diffusion on a modified Chapman layer in which the scale height increases at a rate Γ will be to produce a thinner layer without completely distorting the shape. That is the Yonezawa layer can be quite well represented by a modified Chapman layer with a value of Γ which is higher than that of the Yonezawa layer. Thus there is no serious conflict between the value for the increase in scale height with altitude observed in the backscatter results (0.3) and that inferred from the shape of the equivalent Yonezawa layer (0.125). In other words, in a layer governed by diffusion, the increase of scale height with altitude dH_i/dh of the ions appears to be higher than that of the ionizable constituent dH/dh. Thus the temperature rate of + 2°/km inferred in these results is certain to be an overestimate, possibly by a factor of 2. Wright[90] reports from an analysis of ionosonde data that over the region 250-400 Kms dH/dh = 0.2, and this result is in quite good agreement with the conclusions reached above.

Whereas Yonezawa's theory takes into account variation of temperature with altitude, it assumes that the constituents above 180 kms are unmixed

and distributed according to their respective scale heights. An alternative version of the basic theory (put forward by the Cambridge group[77]) suggests that the constituents are mixed quite well at least up to the level of the F_2 maximum, and to a lesser extent beyond. Thus in this theory the maximum electron density is controlled by two diffusion processes: (1) diffusion upwards of O_2 which controls the loss rate, and (2) downward diffusion of the ions in their attempt to seek hydrostatic equilibrium, and also a third effect—that of large-scale mixing which assists process (1). This third effect can completely upset the distribution which would otherwise result from the first two mentioned processes. Ratcliffe,[77] and Rishbeth and Setty[78] have proposed that the temperate latitude winter anomaly may then be the result of a change in the degree of mixing. Gliddon and Kendall[33] have shown that the Yonezawa layer will satisfactorily explain the diurnal variation of N_{max} and h_{max} for the F_2 region, but it is clear that the theory incorporates no features which would explain the winter anomaly. The Cambridge theory as developed by Rishbeth and Setty[78] is attractive in this respect. Ratcliffe[77] has shown that the ionosonde data can best be interpreted as indicating that the scale height of the ionizable constituent H near the peak of the F_2 region is 50 kms. This is certainly in agreement with the results of the Faraday rotation experiments, and is close to the satellite value for the scale height H_N of the neutral particles at this altitude. However, Ratcliffe supposes that the mean molecular mass of the neutral particles corresponds to N_2 and is therefore 1.75 times that of the ions (O). This then forms the basis of the arguments which suggest that large scale mixing must be taking place. Thus Ratcliffe concludes the scale height H of the ionizable constituent (O) is approximately the same as that H_B of the neutral particles (O_2) which are responsible for the attachment process (i.e., $H \approx H_B$). Ratcliffe[77] argues on the basis of the early backscatter results of Bowles[12] that the ratio $H/H_B \approx 1.4$ in the region above the maximum and hence the mixing is less perfect at these heights. However if the same arguments are now applied to the present results it must be concluded that $H/H_B \leq 1$ and hence mixing is perfect over the range 400 to 700 Kms. Further there is no clear indication that the mixing becomes less than complete even at the uppermost heights. This conclusion seems difficult to accept, and as it is based upon the assumption that the mean molecular mass at 300 Kms corresponds to N_2 one is inclined to question the validity of this initial assumption. Johnson[45] estimates that the number densities of N_2 and O are equal at about 250 Kms height. It seems probable therefore that at 300 Kms the mean molecular mass is less than that corresponding to N_2 and hence mixing is not so complete as concluded by Ratcliffe. At 700 Kms altitude it seems likely that the atmosphere consists principally of atomic oxygen[45] and then if there were no mixing $H_N = 2H_B$. If the values of H_N (from satellite results) were known more reliably the ratio H/H_B could be obtained and the self-consistency of assumptions that either (a) there is no mixing or (b) there is perfect mixing could be examined. Unfortunately the $\pm$ 30% error in the satellite values of H_N combined with a probable error of about $\pm$ 10% in the backscatter results of H preclude these tests at present.

Thus to summarize this section we may say that arguments which suggest that the mixing is nearly complete at 300 Kms altitude are suspect because if these are accepted the present results indicate that mixing is also complete up to 700 Kms altitude. On the other hand the mixing may be less than complete at the level of the maximum and beyond if the assumption that the mean molecular mass corresponds to N_2 at all these heights is abandoned.

VII. Conclusion

The principal results of Faraday rotation experiments and radar backscatter measurements have been summarized. While it is clear that these two new techniques have not yet furnished a wealth of results comparable with those of ionosondes, it is obvious that these techniques when employed in conjunction with ionospheric soundings can provide a great deal of information not available from soundings alone. The results so far support the theory for the F_2 region advanced by Yonezawa and appear to be in slight conflict with a variation of this theory proposed by the Cambridge group. It is shown that a small modification of the theory will resolve the conflict and thus yield a hypothesis which is capable of accounting for (a) the diurnal variation of the F_2 region, (b) the geomagnetic anomaly, (c) the temperate latitude winter anomaly.

References

1. Aitchison, G. J., and Weekes, K., J. Atmos. Terr. Phys, *14*, 236, 1959.
2. Al'pert, L. Ia., Bobriakova, F. F., Chudesenko, E. F., and Shapiro, B. S., Doklady AN USSR, *120*, 743, 1958.
3. Appleton, E. V. and Barnett, M. A. F., Nature, *115*, 33, 1925. See also Proc. Roy. Soc., A *109*, 621, 1925.
4. Bates, D. R. and Massey, M. S. W., Proc. Roy. Soc. A, *187*, 261, 1946. See also, A *192*, 1, 1947.
5. Bauer, S. J. and Daniels, F. B., J. Geophys. Res., *63*, 439, 1958.
6. Bauer, S. J. and Daniels, F. B., J. Geophys. Res., *64*, 1371, 1959.
7. Berning, W. W., J. Meteorol., *8*, 175, 1951.
8. Berning, W. W., J. Geophys. Res., *65*, 2589, 1960.
9. Blackband, W. T., Burgess, B., Jones, I. L., and Lawson, G. J., Nature, *183*, 1172, 1959.
10. Blackband, W. T., J. Geophys. Res., *65*, 1987, 1960.
11. Bowles, K. L., Phys. Rev. Letters, *1*, 454, 1958.
12. Bowles, K. L., Jour. of Res. (N.B.S.) *65 D*, 1, 1961.
13. Bradbury, N. E. Terr. Mag., *43*, 55, 1938.
14. Bramley, E. N., and Ross, W. Proc. Roy. Soc. A *207*, 251, 1951.
15. Bramley, E. N., Proc. Roy. Soc., A *220*, 39, 1953.
16. Breit, G., and Tuve, M., Phys. Rev. *28*, 554, 1926.
17. Browne, I. C., Evans, J. V., Hargreaves, J. K., and Murray, W. A. S., Proc. Phys. Soc. B. *69*, 901, 1956.
18. Browne, I. C., and Evans, J. V., Jodrell Bank Annals *1*, 255, 1958.
19. Chapman, S., Proc. Phys. Soc. *43*, 26 1931.
20. Dalgarno, A., J. Atmos. Terr. Phys. *12*, 219, 1958.
21. De Witt, J. H., and Stodola, E. K., Proc. I.R.E. *37*, 229, 1949.
22. Dougherty, J. P. and Farley, D. T., Proc. Roy. Soc. A *259*, 79, 1960.
23. Dyce, R. B., J. Geophys. Res. *65*, 2617, 1960.

24. Evans, J. V., Proc. Phys. Soc. B. *69*, 953, 1956.
25. Evans, J. V., Proc. Phys. Soc. B. *70*, 1105, 1957.
26. Evans, J. V., J. Atmos. Terr. Phys. *11*, 259, 1957.
27. Evans, J. V., and Taylor, G. N., Proc. Roy. Soc., in publication, 1961.
28. Fejer, J. A., Can. J. Phys. *38*, 1114, 1960.
29. Fejer, J. A., Can. J. Phys. *39*, 716, 1961.
30. Ferraro, V. C. A., Terr. Mag. *43*, 55, 1938.
31. Ferraro, V. C. A., An. de Geophys. *17*, 82, 1961.
32. Garriott, O. K., J. Geophys. Res. *65*, 1139 and 1151, 1960.
33. Gliddon, J. E. C., and Kendall, P. C., J. Atmos. Terr. Phys. *18*, 48, 1960.
34. Gordon, W. E., Proc. I. R. E., *46*, 1824, 1958.
35. Gringauz, K. I., Doklady AN USSR, *120*, 1234, 1958.
36. Hagfors, T., J. Geophys. Res., *66*, 1699, 1961.
37. Hame, T. G., and Stuart, W. D., Proc. I.R.E., *48*, 364, 1960.
38. Hame, T. G., and Stuart, W. D., Proc. I.R.E., *48*, 1786, 1960.
39. Hanson, W. B., and McKibbin, D. D., J. Geophys. Res. *66*, 1667, 1961.
40. Heaviside, O., Ency. Brit., (Tenth Ed.) *33*, 213, 1902.
41. Heisler, L. H., and Whitehead, J. D., J. Geophys. Res. *65*, 2767, 1960.
42. Hill, R. A., and Dyce, R. B., J. Geophys. Res. *65*, 173, 1960.
43. Hulburt, E. O., Phys. Rev. *31*, 1018, 1928.
44. Jacchia, L. G., J. Geophys. Res. *65*, 2775, 1960.
45. Johnson, F. S., J. Geophys. Res. *65*, 2571, 1960.
46. Kallman-Bijl, H. K., J. Geophys. Res. *66*, 187, 1961.
47. Kazantsev, A. N., Romanova, T. S., and Klementenko, Ia., Radiotekh i Elektr., *3*, 1107, 1958.
48. Kennelly, A. E., Elec. World and Eng. *15*, 473, 1902.
49. Kerr, F. J., Shain, C. A., and Higgins, C. S. Nature *163*, 310, 1949.
50. Kerr, F. J. and Shain, C. A., Proc. I.R.E. *39*, 230, 1951.
51. King-Hele, D. G., and Walker, D. M. C., Nature, *186*, 928, 1960.
52. King-Hele, D. G., and Walker, D. M. C., Paper presented to the COSPAR Symposium in Florence, April, 1961.
53. Little, C. G., and Lawrence, R. S., Jour. of Res. (NBS) *64* D, 335, 1960.
54. Martyn, D. F., Phys. Soc. Ionosph. Conf. (Cambridge 1954), 254, 1955.
55. Martyn, D. F., Aust. J. Phys. *9*, 161, 1956.
56. Millman, G. H., Sanders, F. A., and Mather, R. A., J. Geophys. Res. *65*, 2619, 1960.
57. Millman, G. H., Pineo, V. C., and Wright, J. W., Paper presented to the spring U.R.S.I. meeting, Washington, 1961.
58. Minzner, R. A., Champion, K. S. W., and Pond, H. L., AFCRC-TR.-59-267, 1959.
59. Mofensen, J., Electronics, *19*, 92, 1946.
60. Munro, G. H., and Heisler, L. H., Aust. J. Phys. *9*, 343, 1956.
61. Munro, G. H., Aust. J. Phys. *11*, 91, 1958.
62. Murray, W. A. S., and Hargreaves, J. K., Nature, *173*, 944, 1954.
63. Nicolet, M., and Mange, P., J. Geophys. Res., *59*, 15, 1954.
64. Nicolet, M., "Space Research", Proceedings of the First International Space Science Symposium, Nice, 1960. (Amsterdam, North Holland Publishing Co.) p. 46, 1960.
65. Nicolet, M., Planet. Space. Science *5*, 2571, 1961.
66. Nisbet, J. S., J. Geophys. Res. *65*, 2597, 1960.
67. Nisbet, J. S., and Bowhill, S. A., J. Geophys. Res. *65*, 3601, 1960.
68. Pineo, V. C., Kraft, L. G., and Briscoe, H. W., J. Geophys. Res. *65*, 1620, 1960.
69. Pineo, V. C., Kraft, L. G., and Briscoe, H. W., J. Geophys. Res. *65*, 2629, 1960.
70. Pineo, V. C., and Briscoe, H. W., J. Geophys. Res. in publication 1961. (See also, Lincoln Laboratory M.I.T. group report 30G-0013)
71. Pineo, V. C., Kraft, L. G., and Briscoe, H. W. Paper presented to the spring U.R.S.I. meeting, Washington, 1961.
72. Priester, W., Martin, H. A., and Kramp, K. Nature, *188*, 200, 1960.
73. Rakshit, M., Ind. J. Phy. *21*, 57, 1947.
74. Ratcliffe, J. A., J. Geophys. Res. *56*, 463, 1951.

75. Ratcliffe, J. A., Schmerling, E. R., Setty, C. S. G. K., and Thomas, J. O., Phil. Trans. Roy. Soc. A *248*, 621, 1954.
76. Ratcliffe, J. A., and Weekes, K., "Physics of the Upper Atmosphere" edited by J. A. Ratcliffe (London, Academic Press) p. 434, 1960.
76. Rishbeth, M., and Setty, C. S. G. K., J. Atmos. Terr. Phys. *20*, 263, 1961.
77. Ratcliffe, J. A., Proc. Phys. Soc. *75*, 2, 1960.
78. Rishbeth, M., and Setty, C. S. G. K., J. Atmos. Terr. Phys. *20*, 263, 1961.
79. Robbins, A. R., and Thomas, J. O., J. Atmos. Terr. Phys. *13*, 131, 1959.
80. Robbins, A. R., and Thomas, J. O., J. Atmos. Terr. Phys, *15*, 102, 1959.
81. Salpeter, E. E., J. Geophys. Res. *65*, 1851, 1960.
82. Salpeter, E. E., Phys. Rev. *120*, 1528, 1960.
83. Stewart, B., Ency. Brit. (Ninth Ed.,) *16*, 181, 1882.
84. Taylor, G. N., Nature *189*, 740, 1961.
85. Thomson, J. H., and Roger, R. S., privately communicated, 1960.
86. Trexler, J. H., Proc. I.R.E. *46*, 286, 1958.
87. Van Zandt, T. E., and Bowles, K. L., J. Geophys. Res. *65*, 2627, 1960.
88. Webb, D. H., Sky and Telescope, *5*, 3, 1946.
89. Wright, J. W., J. Geophys. Res. *65*, 185, 1960.
90. Wright, J. W., J. Geophys. Res. *65*, 2595, 1960.
91. Yeh, K. C., J. Geophys. Res. *65*, 2548, 1960.
92. Yeh, K. C., and Gonzales, V. H., J. Geophys. Res. *65*, 3209, 1960.
93. Yeh, K. C., and Swenson, G. W., J. Geophys. Res. *66*, 1061, 1961.
94. Yonezawa, T., J. Radio Res. Labs. (Japan) *3*, 1, 1956.
95. Yonezawa, T., J. Radio Res. Labs. (Japan) 7, 69, 1960.
96. Yonezawa, T., and Takahashi, H., J. Radio Res. Labs. (Japan) 7, 335, 1960.

RADAR STUDIES OF PLANETARY SURFACES

G. H. PETTENGILL
Lincoln Laboratory, Massachusetts Institute of Technology*

I. Introduction

Any discussion of radar studies of the surfaces of objects in the solar system must begin with the moon. As the earth's nearest neighbor in space, the moon has been studied optically intensively for many hundreds of years. No other heavenly body of appreciable size comes so close to us, and, therefore, no other body allows such a detailed study of its surface. While optical measuring techniques have become very powerful within the last century, and offer unparalleled angular resolution, they must rely on the shadows cast by the sun for study of surface relief. At optical wavelengths, too, the measurement of surface reflectivity is confined to a small region of the spectrum—at most, several octaves.

The proximity of the moon means also that it is, by a considerable margin, the celestial object most accessible to radar investigation. Radars now in existence or shortly to be completed, which are capable of contributing to astronomy, span the entire range of frequencies—some ten octaves—which are able to penetrate the earth's surrounding atmosphere. The information which they are gathering, when combined with that already obtained by optical means, will add importantly to our knowledge of the moon.

Among the very largest radars, several have now succeeded in detecting echoes from the planet Venus, and have, therefore, begun to extend the techniques tested on the moon to more distant planetary objects. Since far less is known of the surface properties of the planets, the radar measurements will be relatively more valuable than in the case of the moon.

II. Some Properties of the Moon as a Radar Target

A. *Path Loss*

In the design of a radar investigation, a basic consideration is the strength of the returned echo which may be expected. A convenient way of expressing the combined losses due to geometry and target reflectivity is to define a "path loss." Path loss, as the name implies, is a measure of the attenuation suffered by a signal between the time it is sent out by the transmitting antenna and the time it appears at the receiving antenna. Equipment parameters, such as power, antenna gains, and receiver sensitivity, are specifically exluded (although operating frequency appears intrinsically).

If one considers a transmitter radiating a unit of power isotropically (i.e., with an antenna gain of unity), then an object of effective (radar) cross

*Operated with support from the U. S. Army, Navy and Air Force.

section, σ, at a distance, R_1, will intercept and re-radiate back to the receiver a fraction of the power:

$$\frac{\sigma}{4\pi R_1^2} \tag{1}$$

The radar cross section, σ, is defined as the intercept area of an assumed isotropic scatterer which would yield the flux actually detected for the specified angles of illumination and observation. It may range from very much larger than the actual physical cross section of the target to very much smaller, and may vary depending upon the target's orientation with respect to the observer.

Considering now that fraction of the re-radiated power which is intercepted by an isotropic receiving antenna at a distance, R_2, one finds:

$$\frac{\text{Power received}}{\text{Power transmitted}} = \frac{\sigma}{4\pi R_1^2} \cdot \frac{A}{4\pi R_2^2}, \tag{2}$$

where A is the effective aperture of the isotropic receiving antenna. The fractional power received, under the conditions stated, is defined as the path loss. It can be shown[15] that the effective aperture of an isotropic antenna is given by:

$$A = \frac{\lambda^2}{4\pi}, \tag{3}$$

where λ is the radio wavelength employed. The final expression for path loss is, then, assuming the usual radar case where $R_1 = R_2$:

$$\text{Path loss} = \frac{\sigma \lambda^2}{(4\pi)^3 R^4}. \tag{4}$$

Putting in the mean lunar orbital radius for R, and an average measured value[7, 25] for σ, one finds that the path loss is approximately $1.62 \times 10^{-26} \lambda^2$. For $\lambda = 1$ meter (which is near the frequencies where the best measurements have been made) this corresponds to a loss of 258 decibels. Note that if the cross section is independent of frequency, the path loss varies as the square of the wavelength. Changes in the distance to the moon over the course of a month cause this figure to vary by about $\pm$ 1 db. The signal level to be expected for any given installation may be easily obtained from the path loss by applying the actual transmitted power, and the gains of the transmitting and receiving antennas.

B. *Echo Delay and Doppler Shift*

Supposing, now, that the radar system is sufficiently powerful to obtain detectable lunar echoes, what distinctive characteristics will these echoes exhibit? The moon moves in a slightly elliptical orbit around the earth, taking one month to complete a circuit. The mean radius of this orbit is 384,393 kilometers with peak excursions around this figure of $\pm$ 5%. Thus

the radar echo is delayed approximately 2.6 seconds from the transmitted signal—an interval which is readily detectable even by sluggish human senses, and which may have an annoying effect on two-way communication.

Since the distance to the moon varies as it travels around the earth, there is a component of doppler shift introduced into the returned signal which has a monthly periodicity. The peak value of this component varies slightly from month to month but usually does not exceed 5 parts in 10^7 of the radio frequency employed (corresponding to 150 cps at the 1 meter reference wavelength). A much larger source of doppler shift in the returned echo results from the diurnal rotation of the earth. The magnitude of this component depends on the observer's latitude (i.e., distance from the earth's axis) and on the position of the moon in the sky and, of course, has a daily periodicity. For an observer on the equator, the doppler shift arising from diurnal motion can be as large as 3 parts in 10^6 (900 cps at 1 meter wavelength) when the moon is on the horizon. The exact calculation of doppler shift and echo delay must take account of parallax, and is, in general, sufficiently tedious and involved that it is left to digital computers. The accuracy with which the lunar orbit is currently known to astronomers is such that only the most precise radar measurements now being planned offer any hope of improving the lunar ephemeris.

C. *Surface Effects*

In discussing path loss, above, no mention was made of the detailed way in which reflection occurs at the lunar surface. When the character of the echo is carefully examined, one finds that the returned waveform differs distinctly from that sent out. The moon is a large object, some 1738 km in radius, and if a substantial fraction of the visible surface is effective in returning energy, some portions of the returned signal will have been measurably delayed with respect to others. The resulting interference between echoes from different ranges resembles the multipath effects observed in terrestrial communications on the shortwave bands and tends to distort the transmitted waveform. To the extent that the scattering of radio waves by the lunar surface is independent of frequency, this multipath distortion will be statistically similar at all frequencies. For the person planning communication by way of the moon, multipath distortion may place a severe upper limit on the information bandwidth available, unless special steps are taken in preparing the information for transmission.

As the moon revolves around the earth it also spins about its own axis, just as the earth does, but at a much slower rate. In fact, this rotation is just sufficient to cancel the average angular motion of the moon in its monthly orbit, thus presenting very nearly the same lunar face to the earth at all times. There are small deviations, however, since the elliptical orbit requires that the angular motion referred to the earth be greater at some times than others, whereas the rotation must proceed at a uniform rate. This phenomenon, known as optical libration, creates a rocking motion of the apparent lunar disc as viewed from the center of the earth over the course of a month.

Added to this motion, which is intrinsic in the lunar orbit, is a parallax libration caused by the diurnal motion of an observer fixed on the earth's surface. Because the earth's diameter is an appreciable fraction of the distance to the moon, an observer can, in effect, peer around the sides of the moon at moonrise and moonset. The net effect of these two librations (plus several others which are not explained here) is to cause an apparent rotation of the moon as viewed by the radar. Since this rotation causes some portions of the lunar surface to have a component of motion towards the radar while other portions move away (see Fig. 1), energy scattered from different portions of the surface will have slightly differing doppler shifts. As a consequence, the spectrum of the returned signal will be broadened, causing the received signal to fade. Since the parallactic component of libration generally dominates, the maximum effect is usually noticed near meridian passage where the change in parallax is largest. Note that the spectral broadening caused by libration (as also any consequent fading) will be directly proportional to operating frequency. As one might suspect, it can be shown[9] that a relationship exists between the distribution of power in range (multipath) and the distribution in frequency. Unless certain assumptions of statistical uniformity are made for the scattering of radio waves by the moon's surface, however, one distribution is not uniquely defined by the other. Section III-B shows how a simultaneous measure of both distributions forms a powerful tool to investigate surface characteristics of the moon.

Finally, one may study the way in which scattering from the lunar surface affects the polarization of the incident radio waves. If the surface were smooth, at least to the extent of having no irregularities in surface structure which are comparable to a wavelength, the returned wave would be com-

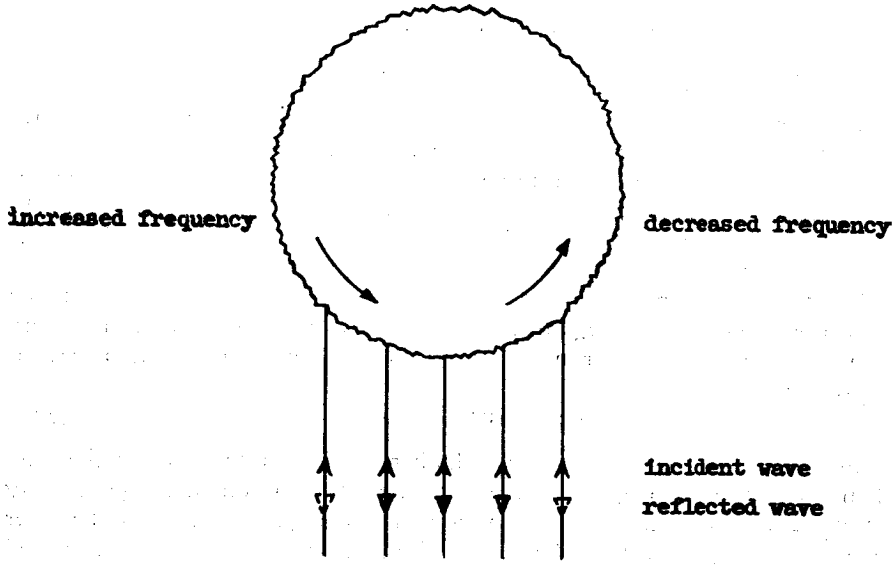


Fig. 1—Diagram showing how a rough, rotating target causes spectral broadening.

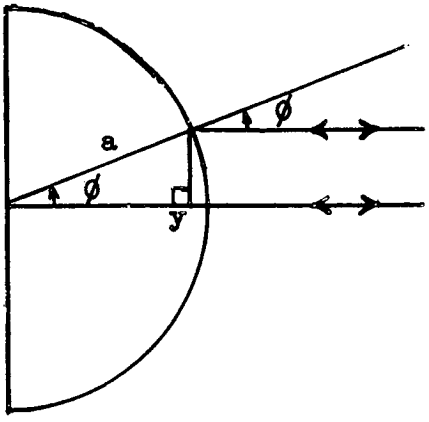


Fig. 2—Geometry for converting range measurements to angle of incidence.

pletely polarized in a way that could easily be predicted from the incident polarization (assuming a knowledge of the Faraday rotation). A rough surface, on the other hand, would cause a partial depolarization of the returned signal. Although surfaces for a symmetrical body can be postulated which would selectively scatter depending on the polarization of the incident wave, the requisite grain structure seems unlikely in the case of the moon. It may be expected, then, that the scattered power will be independent of the incident polarization. Depolarization, of course, will serve as a measure of surface roughness.

III. Status of Current Knowledge of the Lunar Surface

A. *Statistical Properties of the Lunar Surface*

Consider first a short pulse experiment. Figure 3 shows that the returns are smeared in delay (or range). Since the moon is known to be very nearly spherical, one may define an angle, ϕ, measured at its center between the radius to the center of the lunar disc (point of first radar contact) and the radius to the point at which reflection or scattering is considered to take place (see Fig. 2).

Let a be the lunar radius, and y the depth from the point of first contact with the surface to the region in question (the small amount of sphericity which the incident wavefront possesses may be ignored). Then ϕ may be expressed as:

$$\phi = \cos^{-1} (1 - y/a). \tag{5}$$

Now, since ϕ is also the angle at which the incident radiation strikes and then leaves the surface, the observed power distribution in range, y, may be expressed alternatively as a function of angle, ϕ. The measured power level at any delay will be the result of an integration by the transmitted pulse of a small annular region of the surface. This area, for a pulse length,

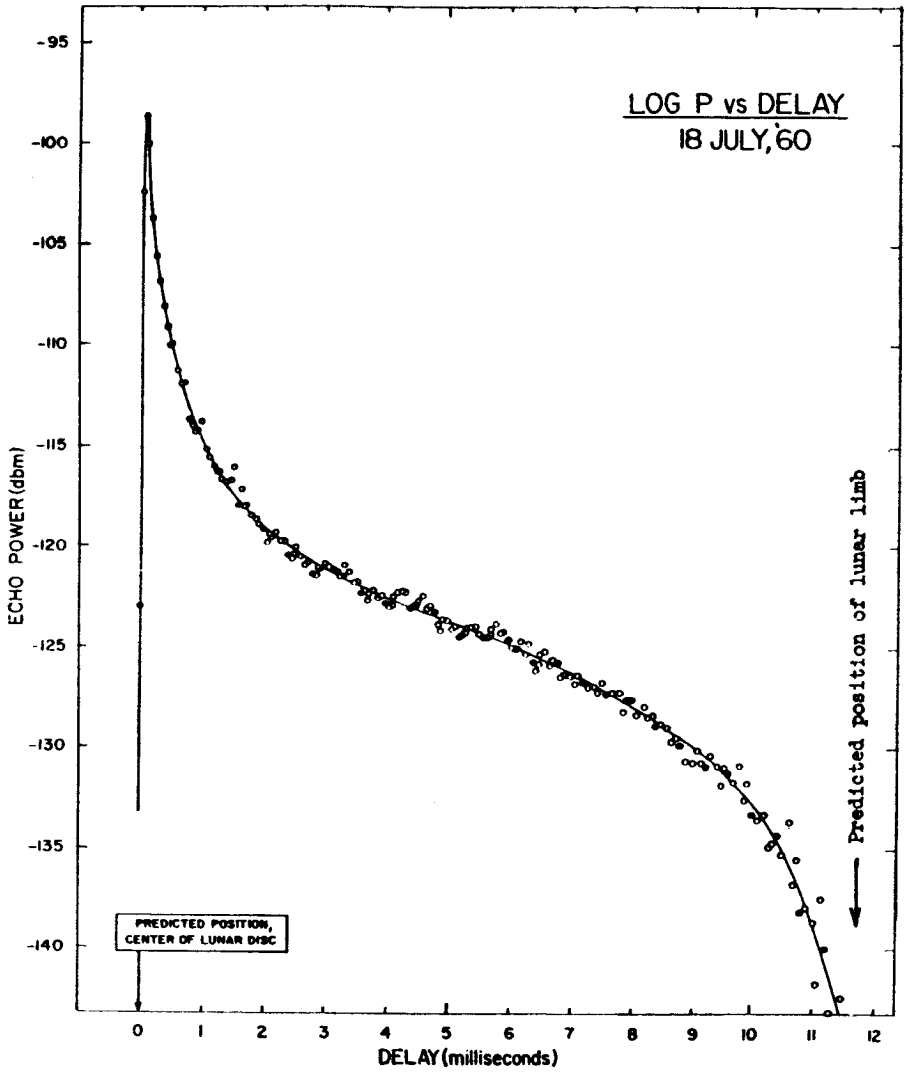


Fig. 3—Plot of received power vs delay for signals reflected from the Moon. These measurements were taken by the Millstone Radar of MIT Lincoln Laboratory using 65-microsecond transmitted pulses and a receiver bandwidth of 35 kcps. Right-circular polarization was transmitted and left-circular received.

τ, and velocity of propagation, c, will be:

$$\Delta A = \pi a c \tau \tag{6}$$

independent of y. Thus one may transform easily to the angular coordinate as shown in Fig. 4. For convenience, since many scattering laws are expressed

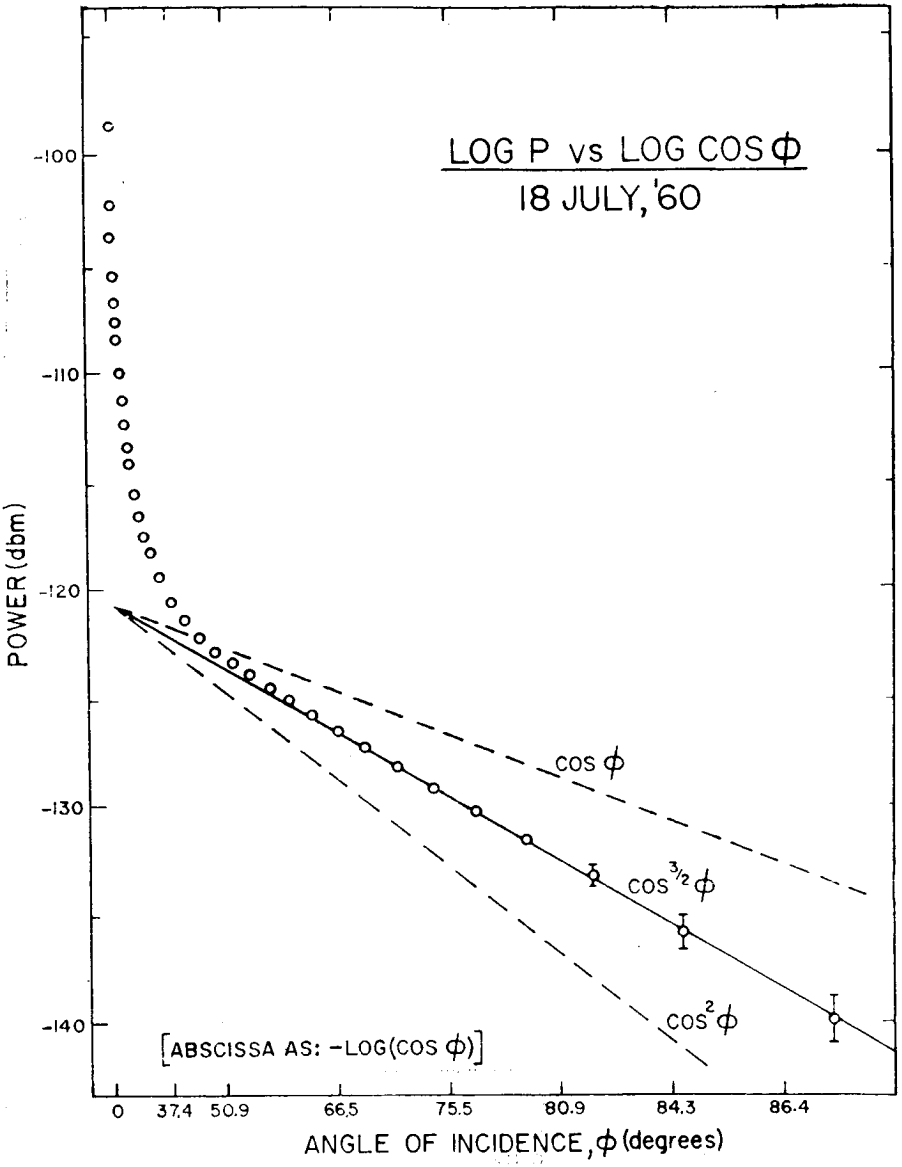


Fig. 4—Plot of received power vs angle of incidence with respect to the lunar surface.

in powers of cos (ϕ), the returned power density has been plotted versus cos (ϕ) using logarithmic scales.

Perhaps the most obvious qualitative feature of Fig. 4 is the presence of two distinctly different regimes. First is seen the rapidly decaying specular-type echo associated with near-normal reflection from the center of the

disc. Then at about 35° (corresponding to a delay of several milliseconds) a weaker but more slowly varying diffuse component becomes visible which persists until the region of the limb is reached.

The sharp echo at the leading edge was first noticed by Trexler[25] and is in marked contrast to the scattering behavior of the lunar surface at visible[17] and infrared[23] wavelengths where no such highlight is noticed. A response so peaked at normal angles of incidence implies, of course, a considerable surface smoothness at radar wavelengths and an absence of steep slopes. Careful measurements by Hey and Hughes[11] and others show that the mean gradient of the smooth portions of the surface lies between 1 in 10 and 1 in 20. It should be remembered, however, that only a comparatively small region of the lunar surface (about 200 km in radius) is sampled by this technique, and that the results may not necessarily apply to the entire surface. Since extremes of the lunar libration cause apparent rotations of as much as 15°, it will be possible to obtain independent measures of the surface smoothness for several regions near the center of the disc.

The diffuse component evident in Figs. 3 and 4 was not discovered until the advent of radars with a very high product of power and aperture because it is so much weaker than the specular echo. In spite of the low power density, however, integration of this component over the entire lunar disc shows that it contributes approximately 20% of the power contained in the total returned signal[21]. The angular behavior of the diffuse component seems to lie between the Lommel-Seeliger and Lambert scattering laws with an angular dependence of scattered power[22] approximately as $\cos^{3/2}(\phi)$. It is interesting to note that this is close to the behavior found by Pettit and Nicholson[23] for scattering in the infrared. By integrating the observed dependence (assuming that it may be extrapolated to $\phi = 0$ without serious error), the total contribution of the component may be found:

$$I_d = P_0' \frac{2a}{c\tau} \int_0^{\pi/2} \cos^{3/2}(\phi) \sin(\phi)\, d\phi = P_0' \frac{4a}{5c\tau} \tag{7}$$

where P_0' is the extrapolated value for power from the diffuse component at $\phi = 0$. A similar integration may be carried out numerically for the specular component and it is found to contribute about 4 times the amount due to the diffuse scattering. The angular dependence of the specular component (when the diffuse contribution has been removed) is plotted in Fig. 5 and appears to be well represented empirically over the entire range of measured values by the expression[22]:

$$P_S = P_0\, e^{-10.5 \sin(\phi)}. \tag{8}$$

A similar result has been obtained by Hughes[12] using 5-microsecond pulses at a wavelength of 10 cm.

A less direct but nevertheless related approach to determining the angular power spectrum has been employed by Evans[5] and Ingalls[13]. By measuring the time auto-correlation function of the signal fading and performing a Fourier transformation on the results, they were able to deduce the mean

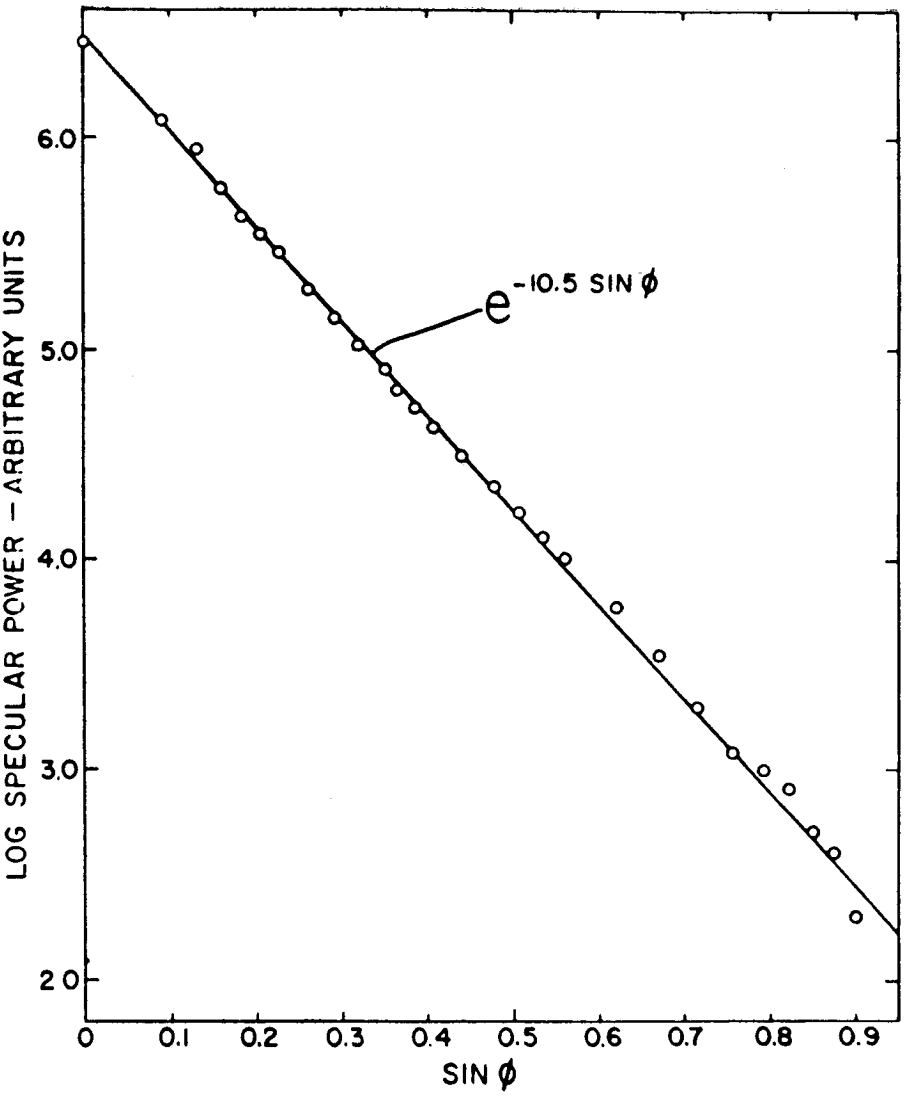


Fig. 5—Angular dependence of the specular component of power scattered from the moon as determined by the Millstone Radar.

normalized frequency spectrum of the scattered signal. Against these measurements may be compared the frequency spectrum derived by Evans[6] from the results of Pettengill[22] and Hughes[12]. The measured angular power spectrum may be converted to a normalized frequency spectrum by a process of numerical integration over the spherical surface (Fig. 10 may help in visualizing the transformation).

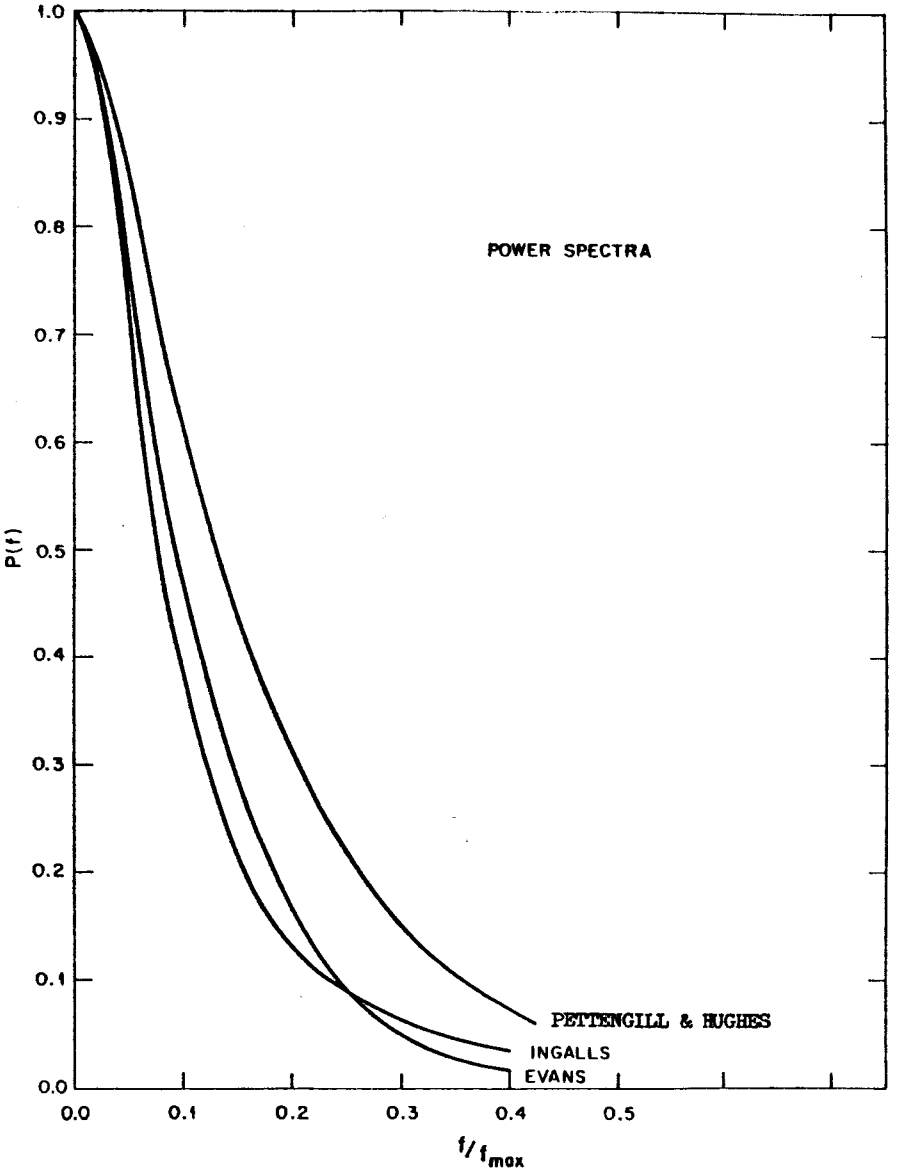


Fig. 6—The power spectra P_f deduced from the autocorrelation results as compared with that derived from the short pulse measurements.

The spectra obtained by the two methods are shown in Fig. 6 where a reasonable agreement may be noted. The slightly broader width of the spectrum derived from the narrow pulse measurements suggests that a departure from the exponential law might be found near the origin (scattering angles less than 3°) were shorter pulses with higher resolution to be used.

A further quantity to which some attention has been directed is the degree of polarization which the scattered signal exhibits. Browne[3] et al reported that the amount of depolarization for plane polarized incident radiation was only of the order of 10%. Using circularly polarized signals, Brown[2] has reported that CW signals with the expected sense at reception were some 15 db stronger than those with the opposite sense. Neither of these determinations made any attempt to distinguish the amount of depolarization associated with echoes from a specific region of the lunar surface.

If the association of the echo "tail" with a diffusely scattering surface is correct, one would expect to find the returned signal significantly depolarized at greater delays as compared to the nearly specular component observed near the beginning. In order to determine this effect, Pettengill[22] has directed a right-circularly polarized transmission at the moon and recorded both right- and left-circularly polarized received radiation. Circular polarization was used since such a mode is not subject to magneto-ionic perturbation in the earth's ionosphere.

In the absence of depolarization, all energy would be returned in the circular sense opposite to that which was transmitted. It may be seen from Fig. 7, however, that while the initial return is highly polarized as expected, there is a substantial amount of energy received in the channel corresponding to depolarization. The polarization of the returned wave may be defined by:

$$P = \frac{I_L - I_R}{I_L + I_R} \tag{9}$$

where I_L and I_R are the scattered intensities for left- and right-circular received polarizations, respectively. Figure 8 plots this quantity for the data shown in Fig. 7. It is clear that a substantial amount of depolarization is associated with the echoes occurring near the lunar limb.

In order to relate the measured powers back to the fractions of the actual lunar surface which may be responsible for a given type of scattering, the directivity of each scattering law must be determined. The directivity factor, g, is a measure of the amount of energy scattered back towards the radar as compared to the average of that scattered into all directions. For the general case, the calculation of g is quite complex since one must taken into account both the phase and amplitude of the scattering from each element of the surface. In the present context, however, with one important exception, the surface elements may be assumed to possess no important phase correlations with their neighbors, and the powers may be simply added.

Under the above assumption, therefore, one may proceed by defining a radar scattering cross section, $\sigma(\phi, \phi', \theta)$ per unit area of the lunar *surface*, where ϕ is the angle between the incident flux and the normal to the surface, ϕ' is a similar angle for the direction in which the scattered energy is measured, and θ is an azimuthal angle locating the plane of the emerging radiation with respect to the incident. Figure 9 shows the geometry. For convenience, $\sigma(\phi)$ is written for the same scattering cross section whenever

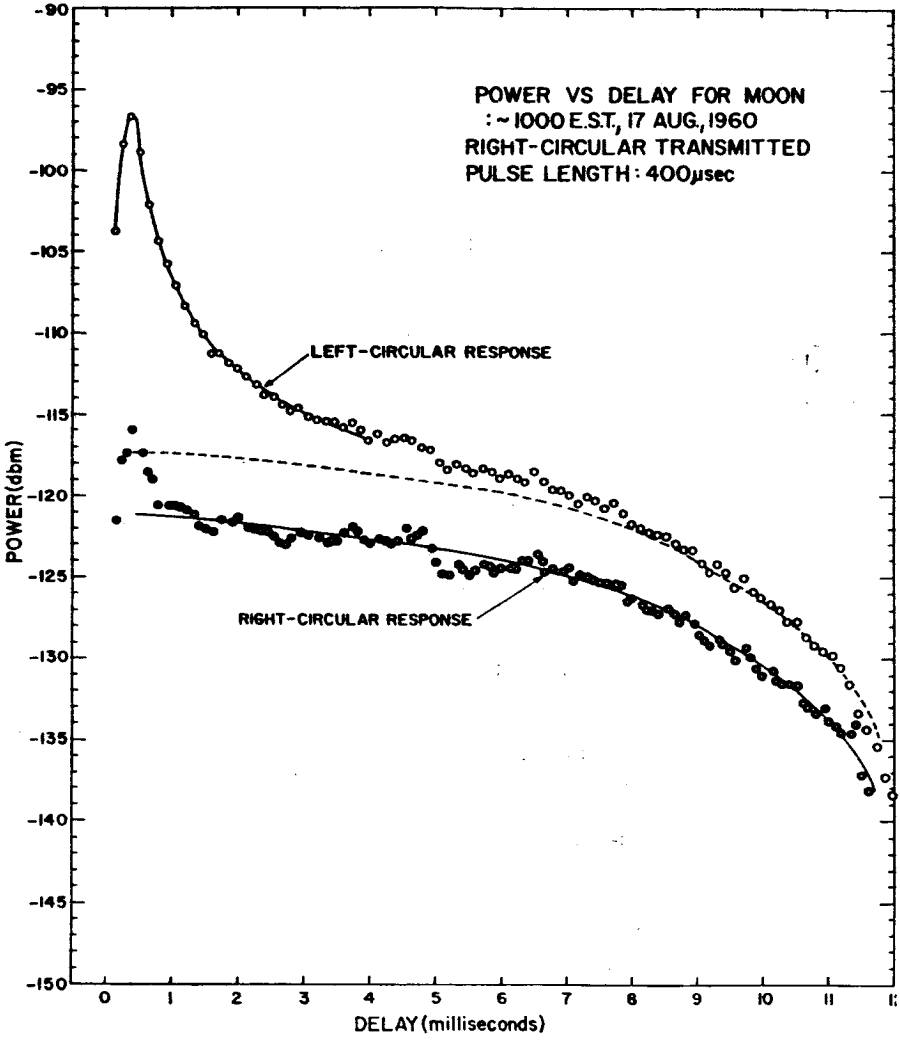


Fig. 7—Lunar echo power vs delay for two orthogonal received polarizations.

$\phi' = \phi$, and $\theta = 0$. One then finds that the directivity is given by:

$$g = \frac{4\pi \int_0^{\pi/2} \sigma(\phi) \sin\phi \, d\phi}{\int_0^{\pi/2} \int_0^{\pi/2} \int_0^{2\pi} \sigma(\phi, \phi', \theta) \sin\phi' \sin\phi \, d\theta \, d\phi' \, d\phi} \tag{10}$$

The generalized cross section, $\sigma(\phi, \phi', \theta)$, cannot be determined by earth-bound radar measurements of the moon, since the available geometries limit ϕ to angles less than 1°. Note that this limitation is not present when the sun serves as the illuminator for measurements in the infrared and optical

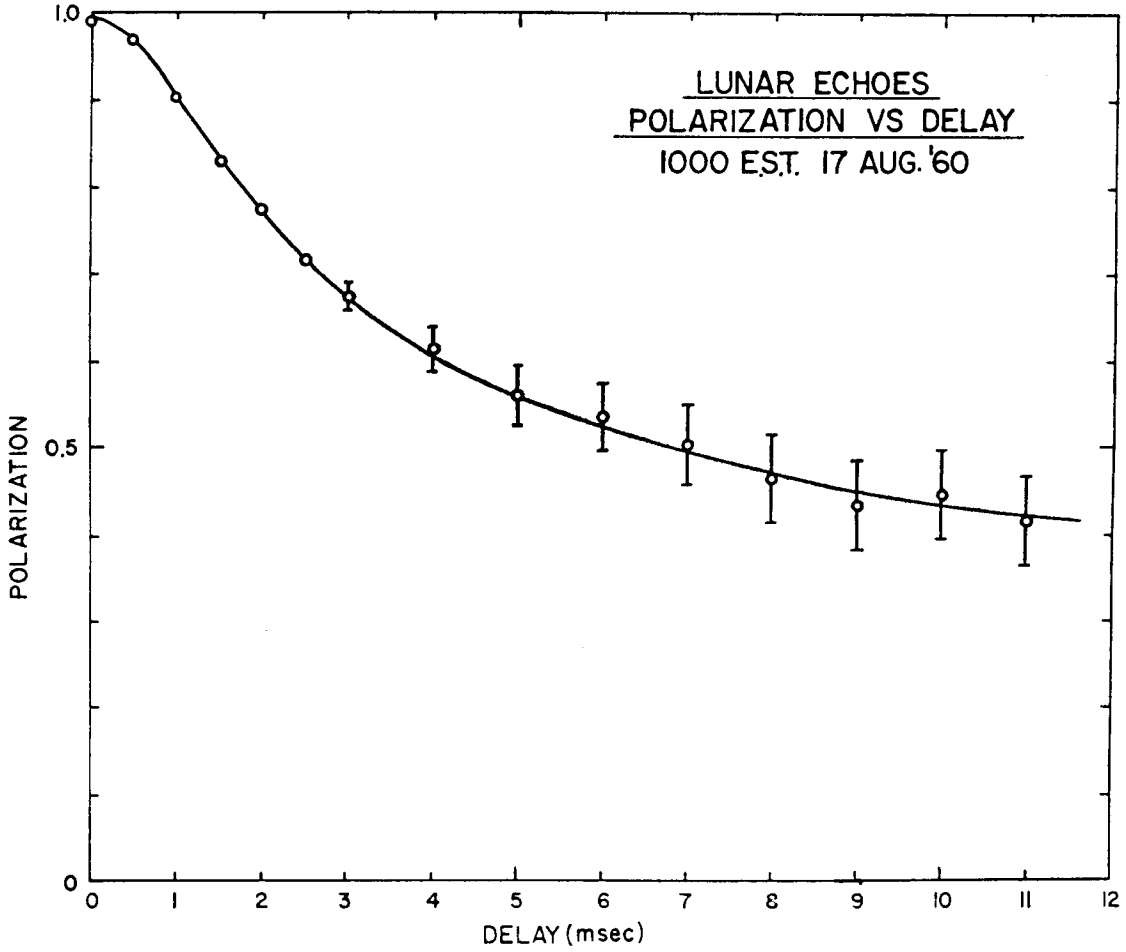


Fig. 8—Polarization vs delay for lunar echoes.

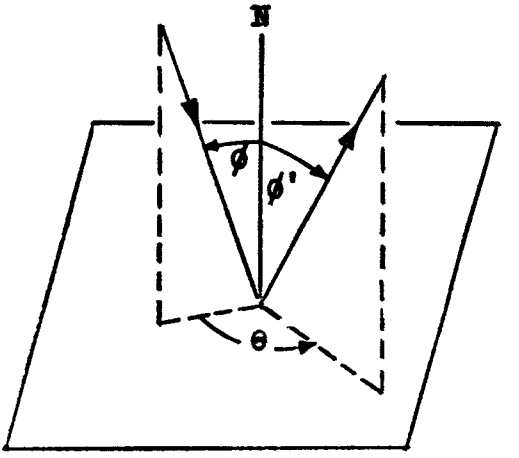


Fig. 9—Geometry for generalized radar scattering cross section.

regions. Despite the radar limitations, however, it is useful to investigate some of the existing scattering laws. For example, Lambert's Law is defined as:

$$\sigma_L\ (\phi,\ \phi') \ \sim\ \cos\phi\ \cos\phi' \tag{11}$$

yielding $g = 8/3$. Similarly, by making infrared measurements over the course of a month, Pettit and Nicholson[23] find (using our definitions):

$$\sigma_{IR}\ (\phi,\ \phi') \sim \frac{\cos\phi'}{(1 + \cos\phi)\ (0.46\ \cos\phi' + \sin\phi')} \tag{12}$$

yielding $g \cong 5/2$. In general, if the scattering can be expressed as:

$$\sigma\ (\phi,\ \phi' \sim \cos^m\phi\ \cos^{n'}\phi, \tag{13}$$

then

$$g = \frac{2(m+1)\ (n+1)}{(m+n+1)}\ . \tag{14}$$

Most of the scattering laws which might apply in the specular region, however, do not fit the dependence given above. A case of extreme importance is the reflection expected from a perfectly smooth sphere (assumed large compared to the wavelength). Since the assumptions of incoherence mentioned above are not valid for this case, formula No. 10 does not apply. A simple calculation using physical optics[20] will show, however, that the corresponding directivity is unity; in fact, the sphere will scatter equal power in all directions. It seems likely that the directivity for nearly specular distributions will, therefore, also be close to unity. Under these assumptions, the rough areas of the lunar surface, which appear to scatter according to Pettit and Nicholson's infrared law, may be assumed to be approximately 2.5 times as effective in backscattering as the smooth portions.

Measurements of the absolute power returned have enabled Trexler[25], Evans[5], and Fricker [7] to calculate the radar cross section of the moon. These measurements are all in fair agreement, spanning the frequency region 100 to 400 Mcps. Fricker's results are probably the most accurate, yielding a CW radar cross section of 0.074 ± 0.01 times the geometrical cross section of the moon. Since it has already been determined that most of the specular energy is returned from near normal incidence to the surface, one may assume a power reflection coefficient, ρ_N, to be associated with the materials near the center of the lunar disc:

$$\rho_N \cong \left(\frac{k^{1/2} - 1}{k^{1/2} + 1} \right)^2 \tag{15}$$

where k is the dielectric constant. The permeability is assumed to be unity and the conductivity zero. The integrated radar cross section for each component may thus set equal to:

$$\sigma = f\ g\ \rho_N\ \pi\ a^2 \tag{16}$$

where f is the corresponding fraction of surface responsible for the scattering. By assuming values for g as stated above and applying the measured ratio of specular and diffuse powers to Fricker's value for the total cross section, it may be shown[22] that only about 9% of the surface is required to be rough in order to explain the observed diffuse scattering. The resulting value for reflection coefficient, taken to be the same for both types of surface, is $\rho_N = 0.064$, yielding $k = 2.81$. Dielectric constants for dry rocks and silicate material of the earth's surface approximate to a value of 5. The apparent disagreement may, perhaps, be traced to a low density of the surface material, at least for those flat areas responsible for the nearly specular reflection. Gold[8] and Van Diggelen[26] have suggested, on both theoretical and observational grounds, that the crater floors may have a brittle, spongy texture of reduced density. Thus, a radio wave incident on these regions would encounter a region of low density with associated low reflectivity. Reflections from denser layers below the surface (if they exist) would suffer absorption in the upper layer and likely would not be seen. Assuming this explanation, one finds a surface porosity of approximately 2 for the flat areas. Note that if the material responsible for the diffuse scattering has a lower porosity (i.e., more compact structure) as seems possible, the estimate of the fraction of the surface responsible for this type of scattering must be revised downwards (changing this fraction will have only a second-order effect on the calculations involving the smooth portions of the surface). In fact, if one assumes a mean density characteristic of terrestrial rocks, a mere 4% of the surface need contribute to the diffuse scattering to explain the observed results.

B. *Mapping Experiment*

Measurements of gross surface properties can yield useful information, as has been shown above, but our understanding of the lunar surface would

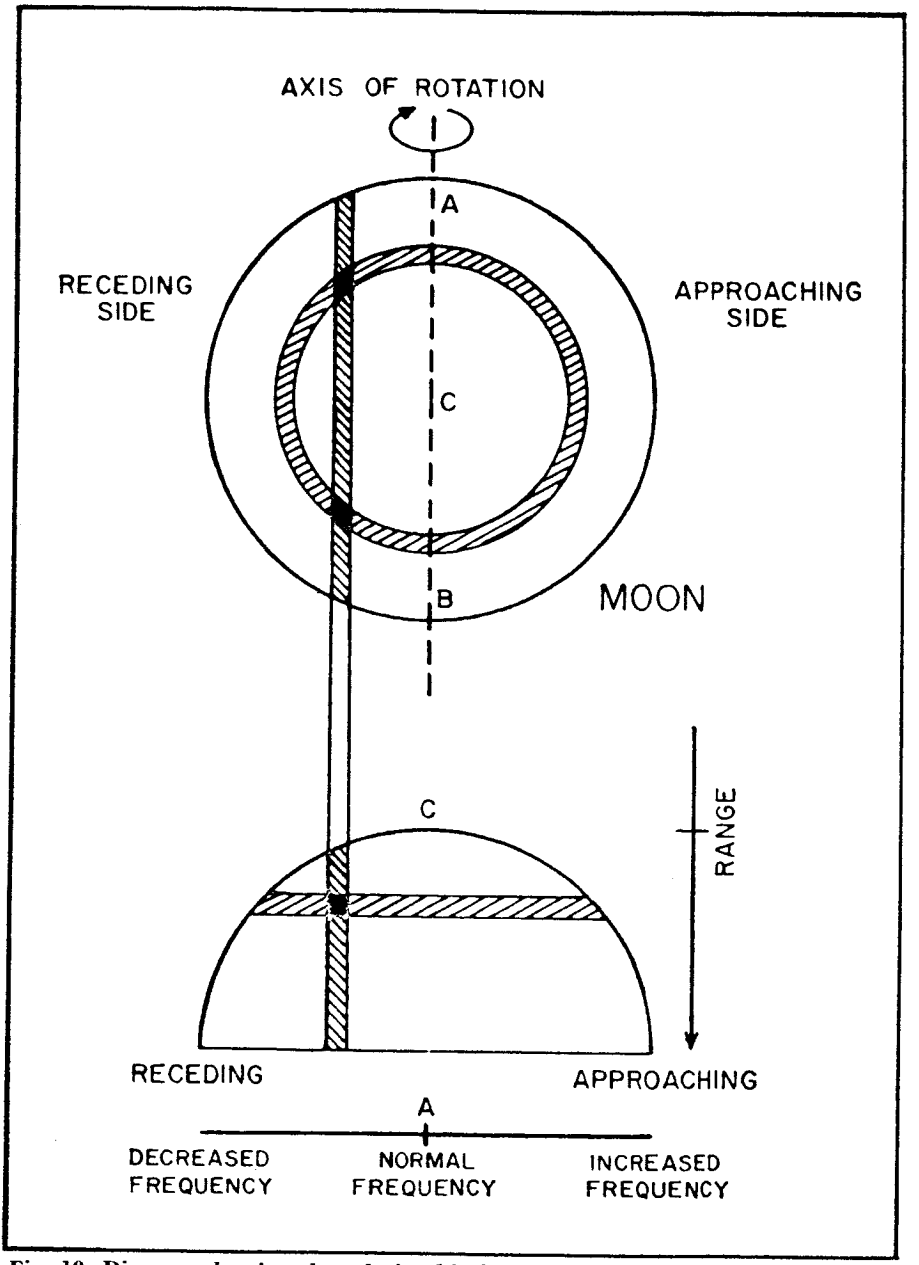


Fig. 10—Diagram showing the relationship between range and frequency contours for two projections of the lunar disc. The upper figure depicts the lunar disc as it is seen by an observer looking along the radar line-of-sight. The lower is a "bottom" view of the illuminated portion of the upper showing how the circles of constant range become thin strips in this projection. The vertical strip corresponds to a contour of constant doppler velocity. Note the way the two contours intersect.

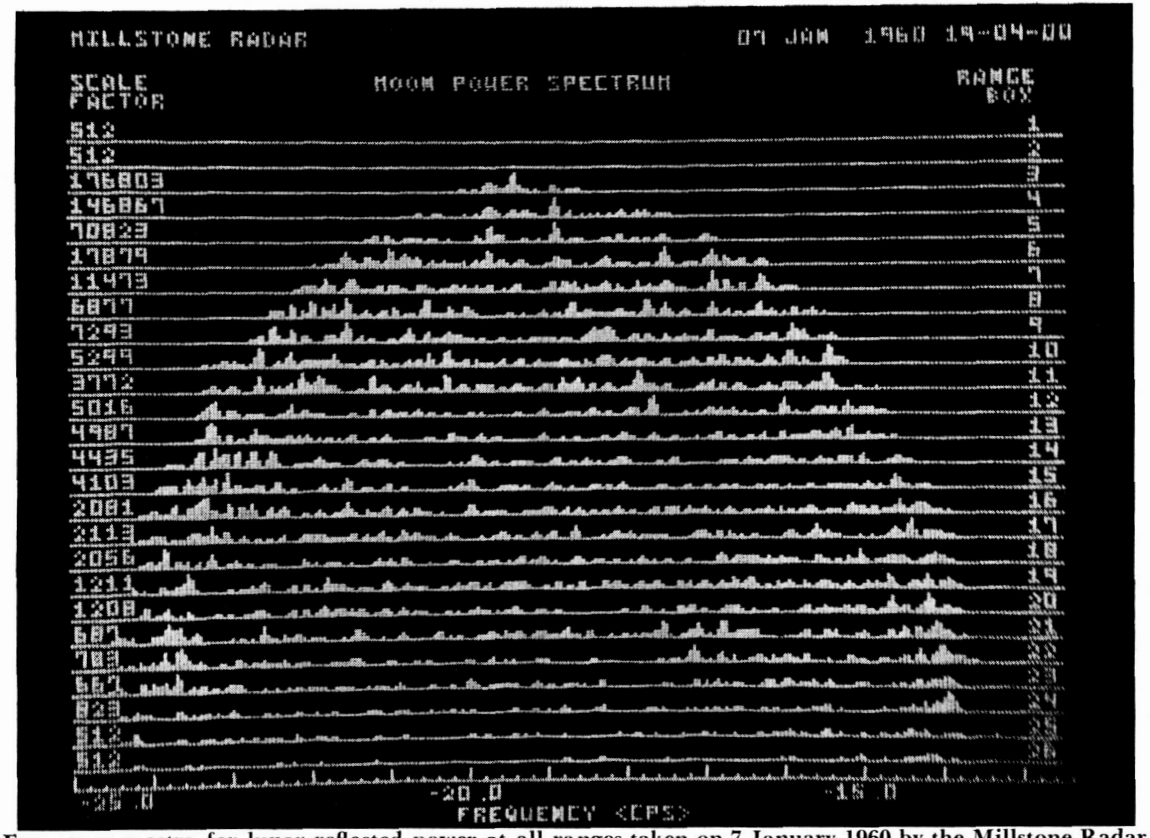


Fig. 11—Frequency spectra for lunar reflected power at all ranges taken on 7 January 1960 by the Millstone Radar. Parameters used are shown in Table I. The display has been generated automatically by the CG 24 digital computer attached to the radar. Note the correspondence with the lower drawing of Fig. 10.

be vastly increased if it were possible to apply the techniques of analysis selectively to small portions of the lunar surface which could be correlated directly with visual observations. Antennas with sufficient angular resolution to accomplish this have not yet been built and will probably not be available for several years. There is, however, another parameter, unique to radar measurements, which can be exploited. This parameter is the frequency shift of the returned signal.

As has been shown earlier (Fig. 1), a rotating body will cause a broadening in the frequency spectrum of the received signal. In fact, each frequency in the broadened spectrum will correspond to a thin strip of the moon's surface oriented parallel to the axis of apparent rotation[14], as shown in Fig. 10. If, now, one were to measure simultaneously the returned power density which lay at a known range and doppler frequency, the scattering source responsible could be located at the intersection of the respective contours. Thus, by carrying out a sufficiently accurate frequency analysis on the power returned at each increment of delay, one would be able to map the surface reflectivity of the moon to within a twofold ambiguity. Note that the frequency shift corresponding to a particular slice of the moon's surface is given by the product of the libration rate and the distance of the slice from the libration axis (added algebraically, of course, to the basic doppler shift associated with the moon's center of mass).

Referring again to Fig. 10, one sees that in the bottom view, where the plane of the paper is taken perpendicular to the axis of apparent rotation, the coordinates (range and doppler frequency) form a cartesian system on which the spherical lunar surface is projected. It is the view that an observer with angular resolution would see if he were able to stand off in space and observe the irradiated portion of the moon from a point at right angles to the radar line-of-sight and to the axis of libration. Because of the twofold ambiguity in intersections, such a plot is actually a superposition of the view from both sides (as if the moon were transparent).

A series of observations of this type have recently been carried out by the Millstone radar of M.I.T. Lincoln Laboratory[21], with an example shown in Fig. 11. Note that the coordinates used correspond to those shown in the bottom half of Fig. 10. Because of the enormous difference between the signal strengths returned from the nose and limbs of the moon (see Fig. 3), the spectra obtained at various ranges have been scaled down by the factors indicated in the left margin. Apart from this adjustment, the height of each element in the small histograms is proportional to power density. The range interval used in these measurements was 500 microseconds (corresponding to a distance of 75 km). Table I shows the radar parameters in use at the time.

At first glance, it may be difficult to understand how one may probe the moon with pulses as short as 500 microseconds and yet preserve the ability to measure frequencies with a resolution of the order of 0.1 cps. The solution lies in the use of a coherent train of short pulses together with sufficient a *priori* astronomical information to resolve the resultant ambiguities. In this case, the interpulse interval must be chosen to be greater than the radar depth of lunar echoes, while the maximum frequency spread associated with

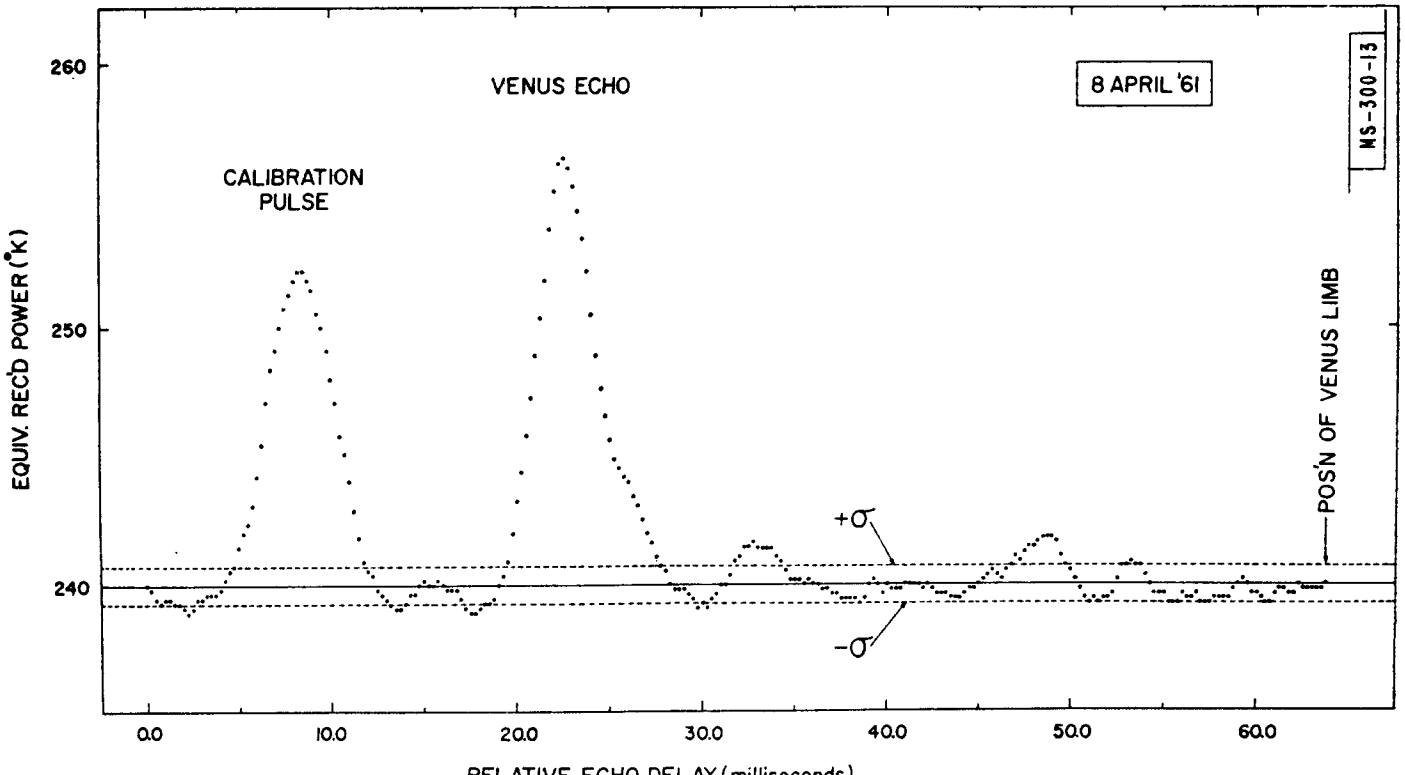


Fig. 12—Venus echo amplitude *vs* delay as seen by the Millstone Radar two days before inferior conjunction in 1961. The transmitted pulse length was 4 milliseconds and the integration extended over 135,000 interpulse periods. The dotted lines show the levels of plus and minus one standard deviation of the noise associated with the measurement.

TABLE I
Radar Parameters

Transmitted frequency:	440.182 Mcps
Antenna half-power (one-way) beamwidth:	2.1 degrees
Antenna gain (above isotropic radiator):	37 db
Overall system noise temperature:	290°K
Polarization transmitted:	right circular
Polarization received:	left circular
Peak transmitted power:	2.1 megawatts
Transmitted pulse length:	500 microseconds
Interpulse period:	29.90 milliseconds
Length of coherent pulse train:	8.97 seconds
Sampling interval:	500 microseconds
Receiver bandwidth:	3 kcps

the libration changes must not exceed the pulse repetition rate. Since the doppler frequency shift is proportional to operating frequency, there will exist a frequency limit above which it will not be possible to apply this method of coding the transmitted waveform for a target of given size and rotation rate. Happily, at 440 Mcps the measurements fall safely below this limit for even the fastest libration rates associated with the moon.

Returning again to Fig. 11, the qualitative features which stand out are: the presence of substantial echoes out to the limb, the sharp bounds to the spectra corresponding to the limits of the moon's surface, and the concentration of signal near the upper and lower frequency limits at the longer delays. The latter effect is geometrical and results from the tangential intersection of range and frequency contours in those areas (see the top drawing in Fig. 10). The measured frequency spread of 10.7 cps at the limb compares well with the predicted value of 10.73. Each run is processed and displayed automatically by the digital computer located at the radar site. It is hoped that the twofold ambiguity which is inherent in this process of mapping the measured power densities back to the surface may be resolved by comparison of results taken at widely separated times. As has been noted earlier, the lunar libration has contributions from several sources. Since these vary at different rates, the axis of apparent libration will incline to the polar axis of the moon at angles which change appreciably (up to 180°) over the course of a day. Therefore, it may be possible to resolve the ambiguity by observing the way in which the echo pattern changes. Up to the present time, however, such an approach has not been attempted.

IV. The Surface Reflectivity and Scattering Law for Venus

The surface characteristics of Venus may be studied by radar in a manner similar to that used for the Moon. The signal strengths obtained from Venus so far are much weaker than for the Moon, of course, and the degree of resolution is limited as a result. Nevertheless, the techniques employed are the same.

Figure 12 shows the result of 2½ hours of integration of the signal received at Millstone where a four-millisecond transmitted pulse at 440 Mcps has been

used[28]. The Venus echo may be seen to rise some 23 standard deviations above the mean noise level. Included in the receiving system was a calibration pulse of known position and intensity which is visible on the trace. From the relative heights of the two signals and using the known radar parameters it is possible to calculate that the radar cross section of the planet is 0.12 $(+ 0.06 - 0.04)$ times the projected geometrical cross section. Similar measurements made at Goldstone[29] indicate a radar cross section at 2388 Mcps lying between 0.10 and 0.15 times the geometrical area, indicating no substantial difference at the higher frequency. It may be seen from Fig. 12 that there is no significant lengthening of the returned pulse, although there is a hint of a delayed component which is just visible. It is, therefore, tempting to apply the same analysis as was used for the Moon, and to deduce that the surface reflectivity is in the neighborhood of 0.12 (again the directivity, g, is set equal to unity). By following through the formula (15) which may be modified:

$$k \cong \left(\frac{1 + \rho_N^{1/2}}{1 - \rho_N^{1/2}}\right)^2 \tag{17}$$

where ρ_N is the reflectivity at normal incidence and k the dielectric constant, one finds that $k \cong 4.25$ $(+ 1.9/- 1.0)$.

The low reflectivity clearly indicates the absence of water as a reflecting surface, although, of course, one cannot on this basis exclude the possibility of large bodies of water distributed over regions which do not include the subterrestrial point. The value for the dielectric constant given above appears to be somewhat higher than that obtained for the Moon ($k \cong 2.81$). It is, however, quite consistent with a value of 5 often quoted for the rocks of the terrestrial surface[30], and does not exclude the possibility that the surface is composed of an ocean of oil as described by Hoyle[31]. If, however, as suggested by Sagan[32] the surface of Venus consists of a layer of oil overlying an ocean of water, then the depth of the oil must be sufficiently thick to absorb the bulk of the energy which would otherwise be reflected from the underlying water surface. The low reflectivity which appears to be essentially independent of frequency, at least in the range 440 to 2388 Mcps, also rules out a dense Venusian ionosphere[33] as the possible reflecting layer responsible for the echoes.

Since very little evidence of pulse lengthening was observed using 4-ms pulses (Fig. 12), shorter pulses were transmitted in an attempt to obtain a measure of the scattering behavior. Figure 13 represents the best results obtained at Millstone (measurements using such short pulse widths were not feasible at Goldstone nor at Jodrell Bank). The signal-to-noise ratio is substantially poorer when short pulses (360 microseconds) are employed, both because the power becomes distributed in range and because the wider receiver bandwidths necessary to accommodate the shorter pulses pass more noise power. Nevertheless, a "tail" to the returned signal is apparent.

Included for comparison in Fig. 13 is a curve obtained by scaling direct experimental measurements of the lunar scattering by the ratio of the radii

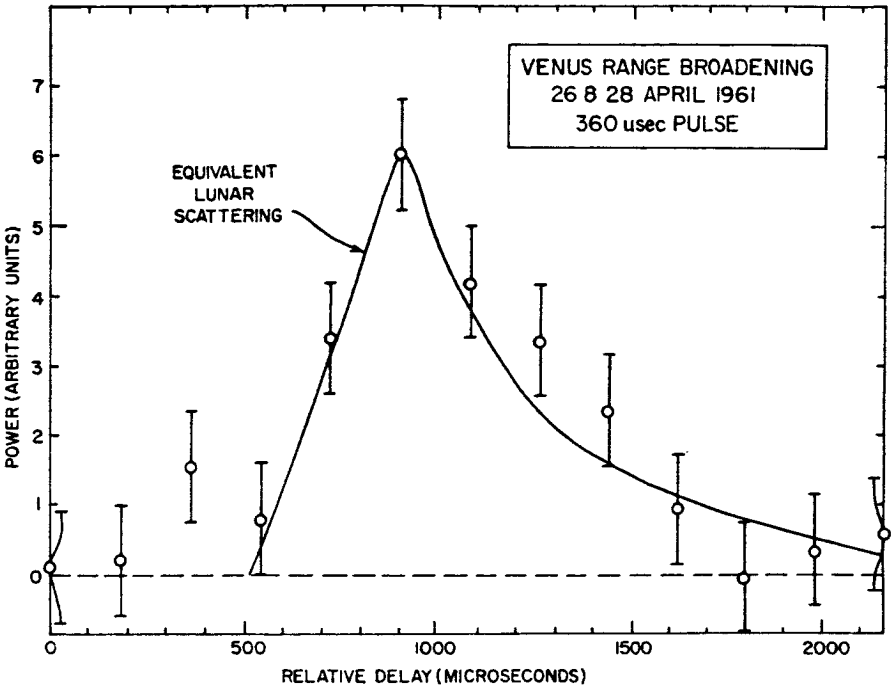


Fig. 13—Venus echo amplitude *vs* delay for 360-microsecond transmitted pulses. A scaled curve showing the angular scattering observed for the Moon is included for comparison.

of the two bodies. Since the radius of the Moon is approximately 1/3.6 times the radius of Venus, a measurement of the Moon using 100-microsecond pulses should yield a distribution of power in range which will have suffered the same convolution throughout as the Venus measurements using 360-microsecond pulses. The results of such a measurement, with the time-scale expanded by 3.6 and the power density normalized to the Venus results are shown in Fig. 13. It may be seen that to within the accuracy of the measurements the agreement is good, indicating that at small scattering angles, at least, Venus has an angular scattering law very similar to the Moon's. Thus, the surface of Venus must be relatively smooth. At present, however, it does not appear possible to discriminate between a solid and a liquid surface, since winds might easily produce sufficient irregularities in the surface of a liquid to explain the observed results.

V. The Planetary Rotation Rate

It has previously been shown how a rough body with a suitably inclined axis of rotation will broaden the spectrum of a signal reflected from the surface. Since the angular scattering distribution has been determined, at least roughly, above, a measurement of the frequency spectrum of the re-

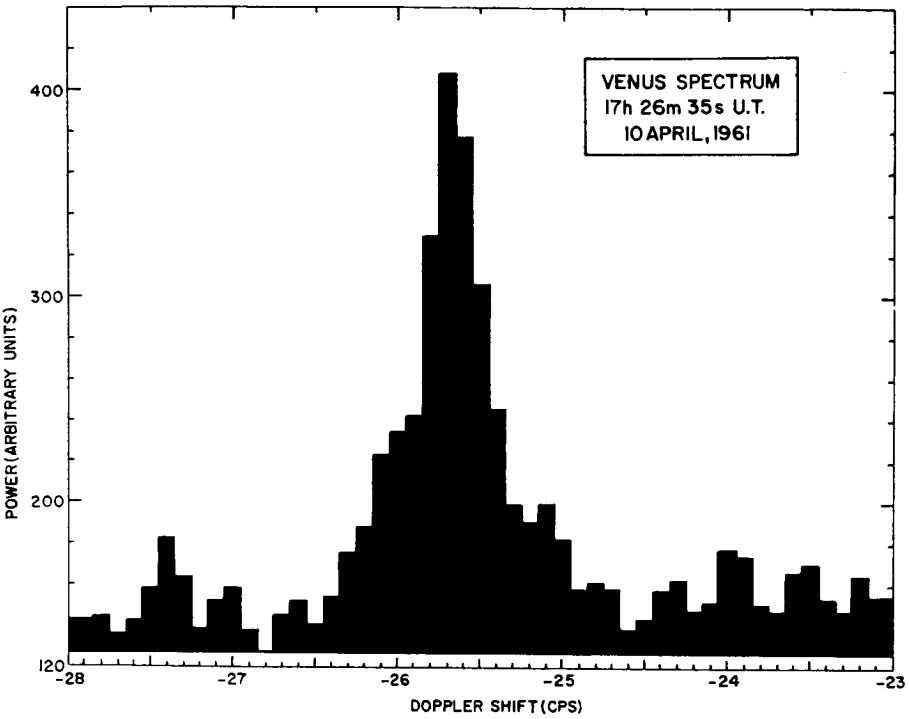


Fig. 14—The spectrum of power returned from Venus within the first 2 milliseconds of the echo duration. The plot represents the mean power density obtained by averaging 120 separate spectral determinations (with all doppler shifts reduced to the planetary reflection time shown). Each spectrum consisted of an untapered Fourier transform of four seconds of data (100 pulses) yielding a resolution of approximately 0.2 cps.

turned signal will yield the component of the planetary rotation lying in the plane at right angles to the radar line-of-sight. Figure 14 shows the spectrum obtained a few minutes after local conjunction on 10 April 1961. It was obtained by averaging 120 separate spectra, each taken with a resolution of approximately 0.2 cps, spread over 14 minutes of time. All spectra were reduced to the absolute doppler measured at the time shown.

It is apparent that the spectrum is extremely narrow, being only about 0.5 cps between the half-power points for a carrier frequency of 440 Mcps. The extreme coherence of the returned signal was first observed with the Goldstone equipment[29], although they find a half-power bandwidth of 9 cps at 2388 Mcps which, when adjusted for the difference in operating frequency, is some 4 times wider than that found by Millstone.

Instead of calculating the amount of spectral broadening which the distribution of Fig. 13 would cause for various rotation rates, it is easier to take advantage of the similarity to lunar scattering. The lunar spectrum was, therefore, measured in a manner similar to that employed on Venus, again

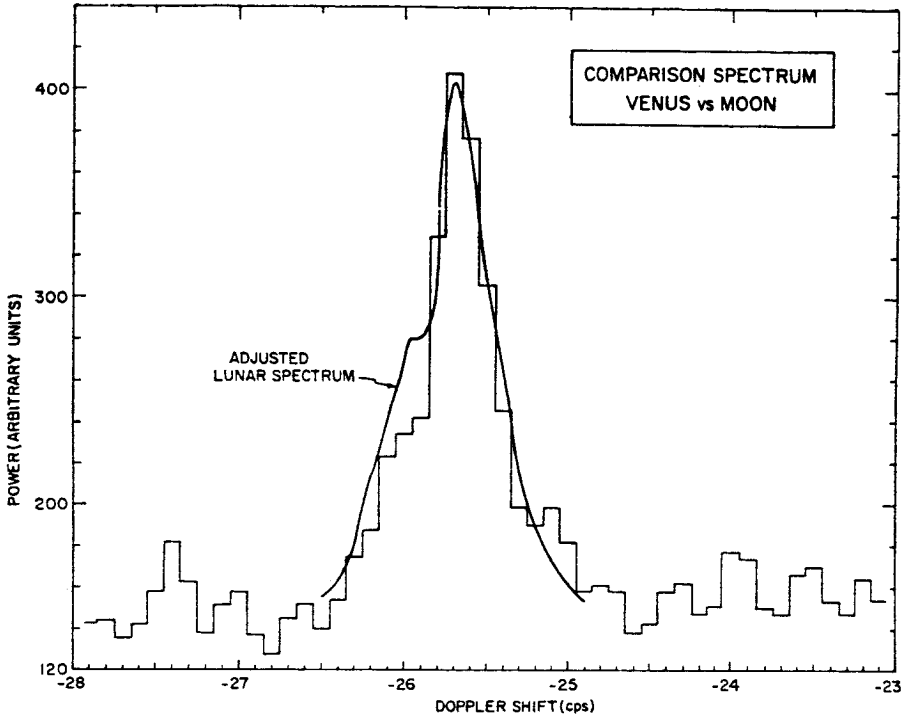


Fig. 15—The spectrum of power returned from Venus compared with a scaled curve obtained from the Moon. The scaling applied to the lunar data is equivalent to an assumption of 6 cps for the limb-to-limb doppler extent of the Venus echo.

using scaled pulse widths. (In this case, the Venus data represent the spectrum of the power returned in the first 2 milliseconds of the return, while for the Moon 500-microsecond pulses were used.) A comparison between the two spectra is given in Fig. 15, where the frequency scale of the lunar spectrum has been adjusted to match that for Venus. Since the lunar libration rate as well as the radii of the two targets was known, it is possible to calculate the full limb-to-limb spectral width which the Venus signal possessed and the associated apparent angular rotation. These are, respectively, 6 cps and 1.7×10^{-7} radians/sec.

The angular rotation of the planet with respect to the radar line-of-sight has contributions from several sources. For most planets in the solar system (except Mercury) the "intrinsic" or sidereal rotation is the dominant. Added to this will be a component arising from the relative geometric positions of the radar and target in their orbits around the Sun. Finally, and least important, is a small contribution from the changing parallax caused by the diurnal rotation of the Earth. For Venus, the intrinsic rotation is unknown but the other factors can be calculated.

For convenience a coordinate system is chosen in which the intrinsic rotation is referenced to the Sun. In this system, a planet such as Mercury which keeps one face always towards the Sun will have zero intrinsic rotation. Since the orbits of Venus and Earth are nearly coplanar (they differ by only 3½ degrees) and are also very nearly circular, a simple geometry suffices to describe the relative motions to an accuracy of several per cent. Figure 16 describes the geometry. If now it is assumed that the distance b and c remain constant and that the angle α changes at a steady rate: $\dot{\alpha}$, then the angle β between the Earth-Venus line and the Venus-Sun line changes as:

$$\dot{\beta} = -\dot{\alpha}\left(\frac{b/c - \cos\alpha}{b/c + c/b - 2\cos\alpha}\right). \tag{18}$$

Inserting b = 149 x 10^6 km., c = 108 x 10^6 km., and $\dot{\alpha}$ = 1.24 x 10^{-7} rad/sec yields:

$$\dot{\beta} = -6.21 \times 10^{-8}\left(\frac{1.383 - \cos\alpha}{1.054 - \cos\alpha}\right). \tag{19}$$

At inferior conjunction ($\alpha = 0$), $\dot{\beta}$ = 4.40 x 10^{-7} rad/sec. The contribution from the parallax of the Earth's rotation is 1.1 x 10^{-8} at conjunction and may be neglected for the present accuracy. If Venus had no intrinsic rotation with respect to the Sun, its entire apparent rotation would be given by Equation (19). If it had intrinsic rotation in the same direction as its orbital motion around the Sun, as all other planets have save Mercury, then the apparent rotation would be greater than given by Equation (19).

In the light of the discussion above, the measured value of 1.7 x 10^{-7} rad/sec. at conjunction seems anomalously low. Taken at face value it would require that the planet be rotating at a sidereal rate less than its period of revolution around the Sun—a conclusion that one would be very reluctant to accept for many reasons. In contrast, the Goldstone spectral measurements indicate an apparent angular rate of approximately 6 x 10^{-7} rad/sec., which within the error of measurement is in good agreement with the assumption that Venus maintains the same face towards the Sun at all times

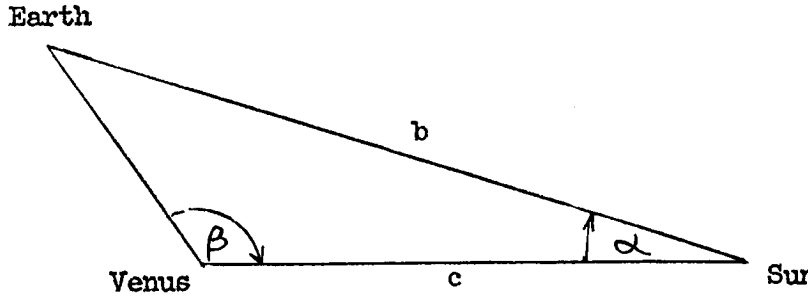


Fig. 16—The Earth-Sun-Venus Geometry.

and has a sidereal period for the rotational component normal to its orbit of 225 days. Note that simple doppler measurements will not reveal rotation about an axis pointed toward the Earth. However, observations made over a period of time will disclose such a component. Therefore, it should be possible to clarify the situation enormously, when all the results of Goldstone and Millstone have been reported. Meanwhile, the available evidence seems to favor a sidereal period of rotation for Venus equal to its solar year of 225 days; i.e., the planet presents the same face towards the Sun at all times. If this conclusion is valid, then observations made at times other than conjunction should display a narrowed spectrum as shown in Fig. 17.

Since it is difficult to find an explanation which can assign the anomalously narrow spectrum obtained by Millstone to experimental error, it perhaps should be borne in mind that the signal-to-noise ratio in Fig. 13 is low and that, possibly, at 440 Mcps the surface is substantially smoother than the Moon, while at 2388 Mcps the surfaces are comparable in roughness. The need for more measurements at both frequencies is clearly indicated.

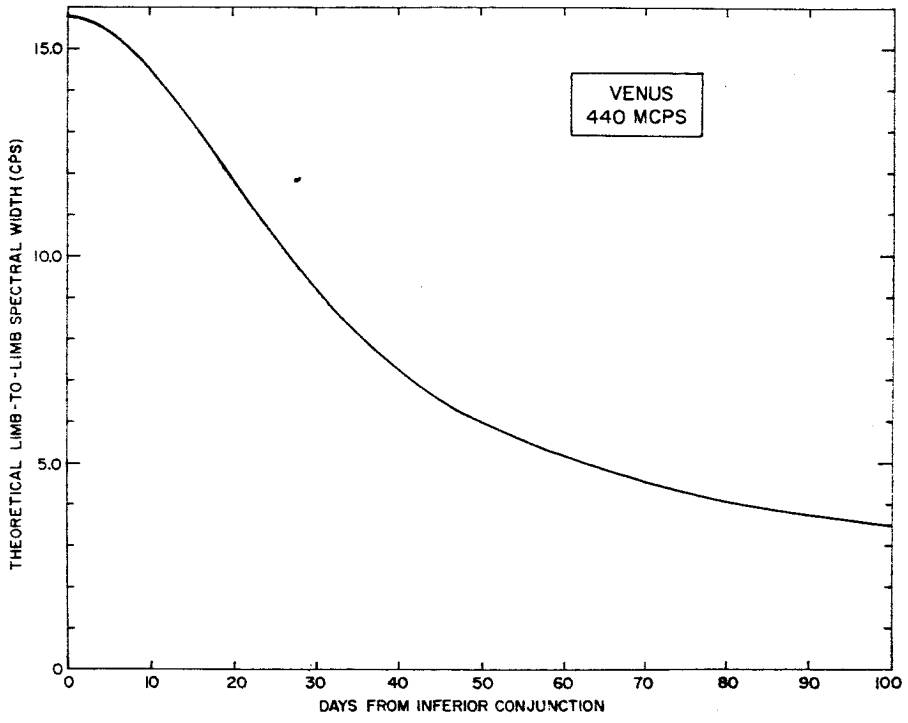


Fig. 17—A plot of the limb-to-limb doppler spread to be expected from Venus at 440 Mcps assuming that Venus moves so as to keep the same face constantly towards the Sun.

VI. Conclusion

It may be seen from the preceding development that a great deal has been learned by the conceptually simple process of echoing radio waves from the surface of the Moon and planets. Knowledge has been gained with every prospect of continued improvement in three major areas: (1) the dynamics of the orbits, (2) the composition of the materials of the surface, and (3) the geography of surface features visible at radio-wavelengths.

In the final analysis, of course, space vehicle landings will be necessary in order to determine with full detail the composition and topography of the various planetary surfaces. Until this explorational era has arrived, however, the information gained by radar will be invaluable, and will find direct application in the design of equipment destined to land on the planets themselves.

References

1. Bay, Z., Hung. Phys. Acta, *1*, p. 1 (1946).
2. Brown, W. E., Jr., Jour. of Geophys. Res., *65*, p. 3087 (1960).
3. Browne, I. C. et al., Proc. Phys. Soc. B., *69*, p. 901 (1956).
4. DeWitt, J. N. and Stodola, E. K., Proc. I.R.E., *37*, p. 229 (1949).
5. Evans, J. V., Proc. Phys. Soc. B, *70*, p. 1105 (1957).
6. Evans, J. V., private communication (1961).
7. Fricker, S. J. et al., N.B.S. Jour. of Res., *64D*, p. 455 (1960).
8. Gold, T., Royal Astr. Soc., monthly notices, *115*, p. 585 (1955).
9. Green, P. E., M.I.T. Lincoln Laboratory Group Report No. 34-84 (Jan. 1960).
10. Grieg, D. D. et al., Proc. I.R.E., *36*, p. 652 (May 1948).
11. Hey, J. S. and Hughes, V. A., "Int'l Astron. Union, Paris Symposium on Radio Astronomy," p. 13 Stanford Univ. Press (Stanford, 1959).
12. Hughes, V. A., Nature, *186*, p. 873 (1960).
13. Ingalls, R. P. et al., Proc. I.R.E., *49*, 631 (1961).
14. Kerr, F. J. and Shain, C. A., Proc. I.R.E., *39*, p. 320 (March 1951).
15. Kraus, J. D., "Antennas," p. 53, McGraw-Hill (New York, 1950).
16. Leadebrand, R. L. et al., Proc. I.R.E., *48*, p. 932 (May 1960).
17. Markov, A. V., Astron. Jour. (USSR), *25*, p. 172 (May 1948).
18. Mitra, S. K., "The Upper Atmosphere," The Asiatic Society (Calcutta 1952).
19. Murray, W. A. S. and Hargreaves, J. K., Nature, *173*, p. 944 (1959).
20. Norton, K. A. and Omberg, A. C., Proc. I.R.E., *35*, p. 18 (1947).
21. Pettengill, G. H., I.R.E., *48*, p. 933 (May 1960).
22. Pettengill, G. H. and Henry, J. C., paper presented to the I.A.U. Leningrad Symposium (No. 14 on the Moon) (1960).
23. Pettit, E. and Nicholson, S. B., Astrophys. Jour. *71*, p. 102 (1930).
24. Sulzer, P. G. et al., Proc. I.R.E., *40*, p. 361 (1952).
25. Trexler, J. H., Proc. I.R.E., *46*, p. 286 (January 1958).
26. Van Diggelen, Utrecht Observatory "Recherches Astron.", *14*, Part II, (1960).
27. Yaplee, B. S. et al., Proc. I.R.E., *46*, p. 293 (January 1958).
28. The Staff, Millstone Radar Observatory, Nature *190*, 592 (1961).
29. Victor, W. K., and Stevens, R., Science, *134*, 46 (1961).
30. Int'l Tel. and Tel. Corp., "Reference Data for Radio Engineers," (New York, American Press, 1956).
31. Hoyle, F., "Frontiers of Astronomy," p. 71 (New York, Harper, 1955).
32. Sagan, C., Science, *133*, 849 (1961).
33. Jones, D. E., Planet. Space Sci., *5*, 166 (1961).

RADAR METHODS OF STUDYING DISTANT PLANETARY SURFACES

JOHN V. EVANS
*Lincoln Laboratory**
Massachusetts Institute of Technology

I. Introduction

In the previous lecture, Dr. Pettengill introduced the general concepts of radar "mapping" of distant planetary surfaces. Deep rough targets distribute power into the time and frequency axes, and by examining this distribution a map can be built up which has a two-fold ambiguity—the reflections from the two halves of the observable hemisphere which lie above and below the apparent position of the equator are inextricably mixed. This distribution of the echo power as a function both of range (or time τ) and doppler broadening (f) has been termed the *target scattering function* $P(f,\tau)$. This function is shown in Figure 1 for three hypothetical spheres, (a) non-rotating, (b) rotating at a rate $W1$, and (c) rotating at a rate $W2$ where $W2 > W1$. Any instantaneous measure of the shape of the function $P(f,\tau)$ would almost certainly appear more irregular than those drawn in Figure 1. This is a consequence of the fact that a radar transmitter provides a source of coherent radiation, and hence the phase of each of the reflected signals is important. At any instant of time the scatterers on a particular part of the surface may reflect largely in phase (constructive interference), or they may cancel one another (destructive interference). However, as the planet rotates the relative path lengths between the transmitter and the scatterers will change, and as a result the signals from any part of the surface will be found to fade. By taking a time average over many reflections, a measure for the distribution of echo power with range and frequency can be obtained, and this will be the same as any subsequent determination of the average scattering behaviour. It may be shown that where the characteristics of the surface are statistically uniform—that is to say the surface is of the same material and has the same kind of roughness everywhere—then the target scattering function $P(f,\tau)$ will be symmetrical about the time (τ) axes. This case—the rotating uniform sphere—is of special interest in radar astronomy. There is increasing evidence that the Moon approximates to such a target, and it seems highly likely that Mercury, Venus and Mars will also behave in a similar fashion. In this lecture it is proposed to explore this special case and decide what can be determined about the roughness characteristics of such a surface from radar measurements.

*Operated with support from the U. S. Army, Navy and Air Force.

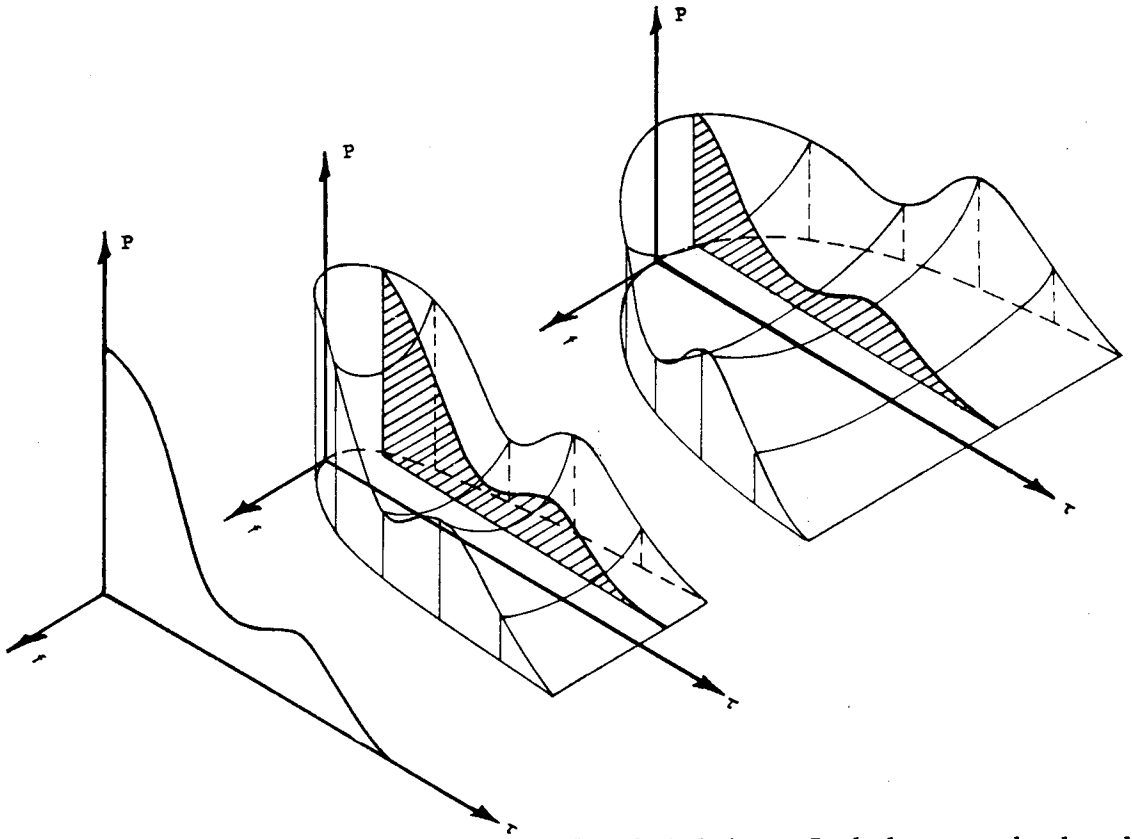


Fig. 1—The *target scattering function* $P(f, \tau)$ shown for three hypothetical planets. In the lower case the planet has no rotational motion and the distribution of the scattering centres can be inferred by measuring the reflected power as a function of range only. In the other cases the planet is rotating (at differing rates) and the echo power is found to be distributed both with time τ (or delay) and with frequency f (or doppler broadening).

II. Experimental Procedures

In this section an outline will be given of the different experimental methods which may be adopted to determine the scattering characteristics of a planet. Most of these methods lead ultimately to the *target angular power spectrum* P_ϕ which describes how the power reflected from an elemental area of the surface varies as a function of the angle ϕ which the incident and reflected rays make with the normal to that element. This function is illustrated in Figure 2. Later in this lecture the relationships between the angular power spectrum P_ϕ and the actual roughness characteristics of the surface will be examined.

IIa. Echo power as a function of range $P(\tau)$

The time average of many echo power *vs.* range displays obtained using a short (ideally an infinitely short) transmitter pulse is called the *infinite impulse response* of the target. Where pulses of finite width are used the display shows the convolution of the pulse shape with the infinite impulse response $P(\tau)$. A time average over many echoes is required to remove the transient local peaks in the power *vs.* range display where, by chance, the reflections from the many scatterers on the surface are largely in phase. The term "scatterer" is perhaps misleading as it implies that certain parts of the surface are more effective reflectors than others, and this would seem contrary to our definition of statistically uniform roughness. There is, however, no contradiction here. Those elements of the surface which are by chance oriented normally to the line-of-sight will produce reflections which will predominate over all others. We shall suppose that such areas are to be found in great numbers and in a wide range of sizes. Thus it is not possible to make a very precise definition of the term "scatterer" without knowing the actual form of the surface roughness.

If the receiver bandwidth is adequate to accept all the returned frequencies (in practice it will always be so for a truly short pulse radar) then the echo power *vs.* range function $P(\tau)$ is the integral over the target scattering function $P(f, \tau)$ with respect to frequency.

$$\mathrm{P}\ (\tau) = \int_{-\mathrm{f_{max}}}^{+\mathrm{f_{max}}} \mathrm{P(f,\tau)}\ \mathrm{df} \qquad (1)$$

where $f_{max} = +\ 2Wa/\lambda$, $W =$ the angular rotation speed (radians/sec), $a =$ the planetary radius (metres), and $\lambda =$ the radio wavelength (metres). The limits of the integral $\pm f_{max}$ are the doppler frequencies (measured relative to the frequency returned from the centre of the disc) of the signals returned from the limbs of the planet. It may be shown that the echo power *vs.* range function $P(\tau)$ can be transformed to provide a measure of the reflected power as a function of the angle of incidence ϕ of the radio waves to the mean surface. This is because there is a unique relationship between the angle ϕ and the range $c\tau/2$ measured from the leading edge of

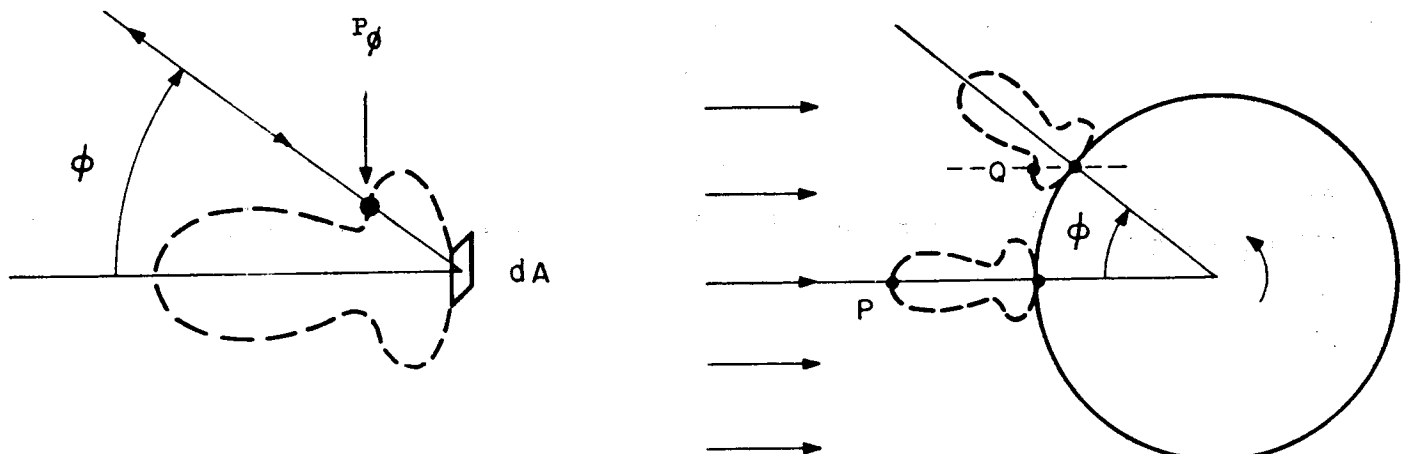


Fig. 2—The manner in which the radar cross-section of an elemental piece of the surface varies as a function of the angle of incidence and reflection ϕ is called the *target angular power spectrum* P_ϕ. This function is illustrated here for a hypothetical piece of surface and the power reflected from the sphere at two points is represented by the length of the horizontal lines through P and Q.

the planet. Both these quantities define an annulus on the surface. By integrating the power reflected from each element of a given annulus the relation between P_ϕ and the echo power as a function of range $P(\tau)$ can be obtained

$$P(\tau)\ dt = \int_s P_\phi\ .\ \cos\ \phi\ da \tag{2}$$

which may be shown to yield

$$P(\tau) = \pi\ acP_\phi\ .\ \cos\ \phi \tag{3}$$

Thus apart from the normalizing constants πac the functions $P(\tau)$ and P_ϕ are related simply by the cosine of the angle of incidence. Measurements of $P(\tau)$ for the Moon have been made by Hughes,[16] Pettengill,[20] and Leadabrand *et al.*[18] from which laws for the angular power spectrum P_ϕ were deduced.

IIb. Echo power as a function of frequency P(f)

For C.W. or very long pulse observations it becomes impossible to measure $P(\tau)$, but instead the doppler broadening of the echo can be found. This is termed the echo power spectrum $P(f)$ and is related to the target scattering function $P(f, \tau)$ in

$$P(f) = \int_0^{\tau_{max}} P(f, \tau)\ d\tau \tag{4}$$

The contours of constant doppler shift are a series of straight lines drawn parallel to the apparent axis of rotation of the planet. This is illustrated in Figure 3. Thus for a planet whose axis of rotation is perpendicular to the line-of-sight from the observer, the distribution of echo-power as a function of frequency is identical to the one-dimensional brightness distribution across the disc. Whilst the brightness distribution over the disc yields immediately the distribution of the scatterers on the disc, it is necessary to solve an integral equation in order to determine P_ϕ. That is, $P(f)$ and P_ϕ are related in

$$P(f)\ df = 2df \int_r^a \frac{P(\sin^{-1} u/a)\ u\ du}{\sqrt{u^2 - r^2}} \tag{5}$$

where $u = a \sin \phi$ {hence $P_\phi = P(\sin^{-1} u/a)$} and $r = f\lambda/2W$ is the distance measured over the projected disc between the apparent axis of rotation and the strip which returns signals shifted in frequency by f. In the case where the axis of rotation is not perpendicular to the line-of-sight, but inclined to it at an angle θ, the rotational motion can be resolved into two vectors $W \cos \theta$ along the line-of-sight and $W \sin \theta$ perpendicular to it. In the above equations W is then replaced by $W \sin \theta$. The rotation about the line-of-sight introduces no changes in the path lengths between the scatterers and the observer and hence introduces no doppler shifts.

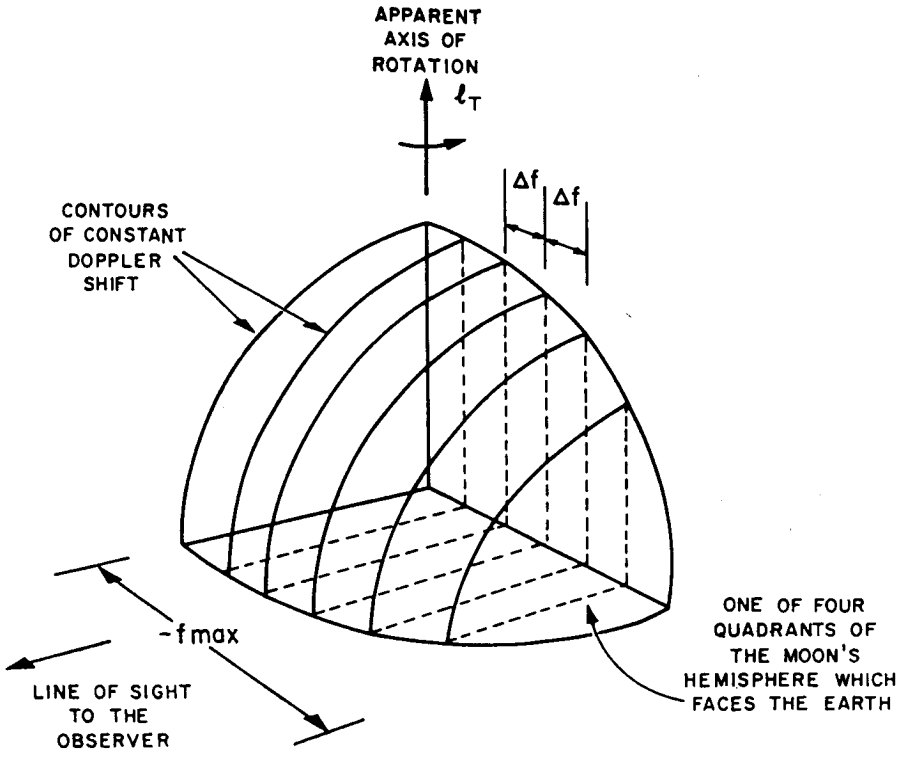


Fig. 3—The contours of constant doppler shift are drawn on the surface of one quadrant of the moon's hemisphere. If the contours are spaced by equal increments Δf, then they map out a set of equi-spaced parallel lines on the projected disc of the moon. The lines are parallel to the axis of librational rotation.

Bracewell[2] has shown that an equation of the form given in Eq. 5 can be inverted to obtain P_ϕ in terms of $P(f)$.

$$P(\sin^{-1} u/a) = \frac{-1}{\pi} \int_u^a \frac{dP(f)/df}{\sqrt{r^2 - u^2}} \, dr \qquad (6)$$

Thus though measurements of the radio frequency power spectrum $P(f)$ can be made to yield the target angular power spectrum the method is not as direct as measurements of the infinite impulse response $P(\tau)$. Nevertheless this method was employed (by Evans[7]) in the earliest attempt to find the angular power spectrum P_ϕ for the Moon. In this work the integral Eq. 5 was not solved using Eq. 6, but instead Eq. 5 was repeatedly integrated for different functions P_ϕ until one was found which gave the correct power spectrum $P(f)$. The power spectrum $P(f)$ for Moon echoes is very narrow ($\sim$ 6cps wide at $\lambda = 1$ metre) and hence Evans was obliged to determine

$P(f)$ by an autocorrelation technique described below (IIc). Fricker *et al.*[11] have attempted to measure $P(f)$ for the Moon directly using a narrow-band receiver as a spectrum analyzer, but encountered many difficulties. The most elegant method which has been employed so far is one due to Pettengill.[20] In this method phase-coherent pulses are transmitted and the received signals are mixed in a pair of phase-detectors with a sine wave having the same center frequency. The outputs of the detectors are the sine and cosine components of the signal, and the amplitudes of these components are converted to digital numbers by an electronic sampling process. These numbers are then stored in a fast digital computer, and later made to yield the radio frequency power spectrum $P(f)$ by instructing the computer to perform a Fourier analysis of the stored phase and amplitude information.

IIc. Echo amplitude autocorrelation measurements

When a planet has a high rotation rate W (e.g. Mars) the radio frequency power spectrum $P(f)$ can be measured directly using a narrow-band receiver as a spectrum analyzer. Where, however, the spectrum is restricted to a few cycles in width (as for the Moon, Venus and probably Mercury) this method is not sufficiently accurate. In this case an alternate procedure is to be recommended. The signals reflected from a rough rotating sphere will be found to exhibit both amplitude and phase fluctuations. If the echo amplitude $y(\tau)$ is measured as a function of time, then the echo amplitude autocorrelation function $\gamma(\tau)$ can be determined from

$$\lambda(\tau) = \frac{\overline{y(\tau)\ y(\tau + \Delta\tau)} - \overline{y(\tau)^2}}{\overline{y^2(\tau)} - \overline{y(\tau)^2}} \tag{7}$$

where the "bars" denote average values.

The Wiener-Kinchin theorem states that the amplitude-autocorrelation function of a fluctuating quantity is the Fourier transform of its power spectrum. Thus by taking the Fourier cosine transform of $\gamma(\tau)$, the *video* power spectrum can be obtained. However, this is not the same as the power spectrum $P(f)$ of the incoming radio frequency signal, because cross-modulation terms are introduced by the detector. Provided that there are more than about ten independent scatterers on the surface, then the received signal will resemble narrow-band Gaussian noise.* For this case the echo-autocorrelation function $\gamma(\tau)$, and the Fourier cosine transform of the radio frequency power spectrum σ are related by

(i) For square-law detectors

$$\gamma(\tau) = \sigma^2 \tag{8a}$$

*The amplitude of the detected signal (using a linear detector) will follow the Rayleigh distribution. This criterion should be applied as a first test in order to determine whether or not the signal resembles narrow-band Gaussian noise.

(ii) For linear detectors the relationship is more complex but approximates to

$$\gamma(\tau) \approx 16/17\ [\sigma^2 + \sigma^4/64] \tag{8b}$$

In practice the echoes reflected by the Moon show very deep fading (30-40 db) and it becomes difficult to build square-law detectors which have a dynamic range as large as this. Consequently most workers have been obliged to use linear detectors.

In the particular case of the Moon, the apparent rotation rate is the sum of three librations, and has both diurnal and monthly variations. The observed autocorrelation function then depends upon the time when the measurements are made. In this case it is convenient to normalize the time τ axis by multiplying by the maximum doppler shift (f_{max}) corresponding to the time of observation. Then all measurements yield a single curve for the autocorrelation $\gamma(f_{max}\tau)$. By taking the Fourier transform of the square root of this, the normalized power spectrum $P(f/f_{max})$ is obtained. One must then solve the integral equation given in (IIb) above to determine the angular power spectrum P_ϕ.

IId. Spaced receiver autocorrelation measurements.

We have stated that the intensity and phase of the signal reflected from a rough rotating sphere will fade. Whilst it is probably most convenient to examine the coherence of the reflected wave-front by making autocorrelation measurements, it is in principle equally appropriate to examine the wave-front by making cross-correlation measurements using spaced receivers. Such measurements constitute one of the most important techniques in radio astronomy where interferometers are used to determine both the angular diameter and precise position of radio sources. The correlation function which defines the phase or amplitude pattern on the ground is simply the Fourier transform of the complex field function which describes the distribution of intensity and phase across the source. Thus spaced receiver cross-correlation measurements determine a spatial correlation function $\gamma(s)$ which Fourier transforms to yield the intensity distribution across the source. This we have seen is the same as the radio frequency power spectrum $P(f)$. For a source which has a statistically uniform surface the cross-correlation function $\gamma(s)$ will be independent of the direction of the base-line.

In methods (IIb) and (IIc), it is helpful to know the rotation rate W of the planet. The principal advantage of the method outlined in (IId) is that it is not necessary to know W to determine the brightness distribution this way. Further, if the cross-correlation function is measured with the receivers at a fixed spacing s, but at different time separations τ then one component of the velocity at which the pattern moves over the ground can be determined. By using three spaced receivers both the rotation rate and the axis of rotation of the planet can be determined. This question has been discussed in some detail by Manasse,[19] who points out that since the rotation of the Earth causes the orientation as well as the effective length of the base-line to change, it is not strictly necessary to employ more than two

stations. An alternative method by which the apparent rotation rate may be determined would be to measure the angular power spectrum P_ϕ using method (IIa) and then compare this with the results of employing either method (IIb) or (IIc).

IIe. Two frequency observations

We have seen that the angular power spectrum P_ϕ may be determined by measuring (i) $P(f)$ the power spectrum or (ii) $\gamma(\tau)$ the echo amplitude autocorrelation function. We note that (i) and (ii) are related by a Fourier transformation. The question then arises "If we cannot determine the infinite impulse response $P(\tau)$ directly can we determine its Fourier transform, and what is this?" T. Hagfors[12] and R. Price[21] have shown that the cross-correlation function $\gamma(\Delta\omega)$ which describes the instantaneous correlation between the amplitudes of two reflected signals having a small frequency separation $\Delta\omega$ is the square of the Fourier cosine transform of $P(\tau)$. That is

$$\gamma(\Delta\omega) = \left[\int_{-\infty}^{+\infty} P(\tau)\ \exp\ (i\tau\Delta\omega)\ d\tau\right]^2 \qquad (9)$$

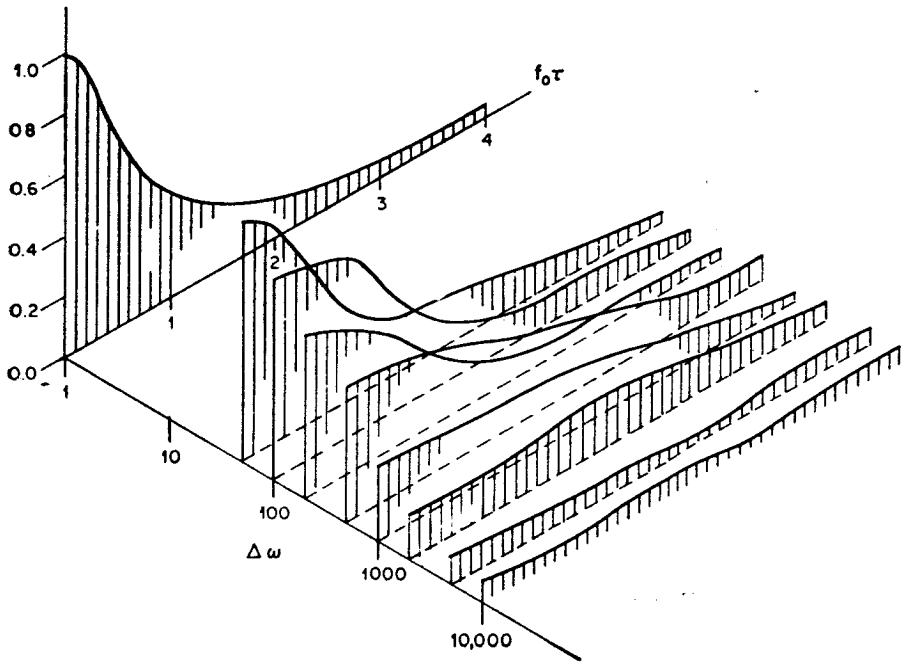


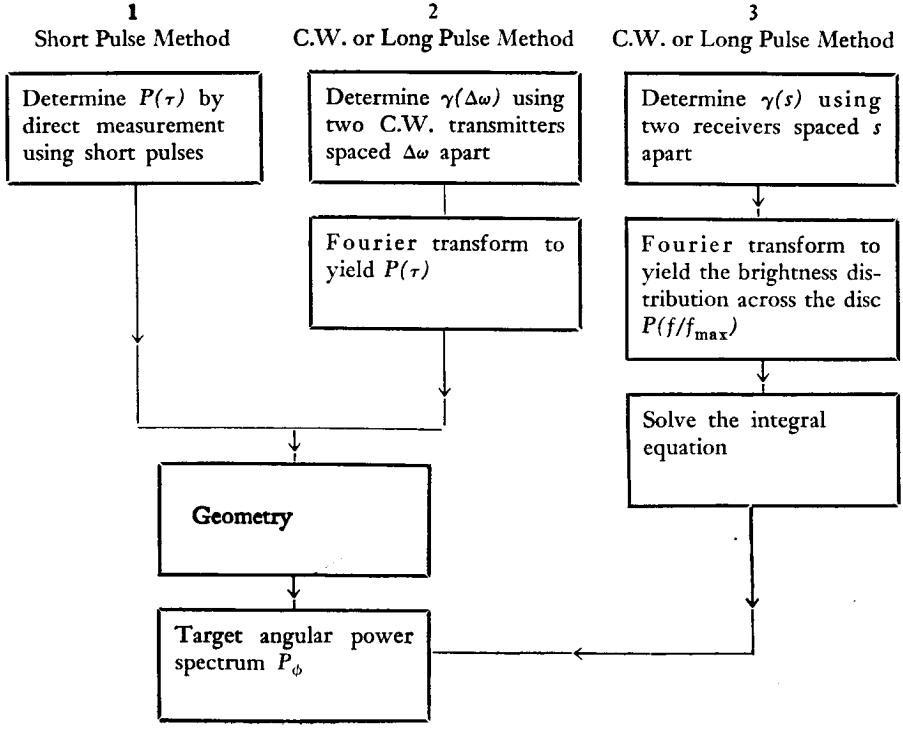
Fig. 4—The correlation of echo amplitude both as a function of τ and frequency $\Delta\omega$ observed by Ingalls et al.[17] In order to remove the effect of the different rates of libration of the Moon, the values of τ have been scaled by the maximum doppler shift ($f_0 = f_{max}$) corresponding to the time of observation. This three dimensional function $\gamma(\Delta\omega, f_{max}\tau)$ is related to the target scattering function $P(f, \tau)$ (Figure 1) by a double Fourier transformation.

Like the echo-amplitude autocorrelation function $\gamma(\tau)$ the two frequency cross-correlation function $\gamma(\Delta\omega)$ has been normalized so that $\gamma(\Delta\omega) \rightarrow 0$ as $\Delta\omega \rightarrow \infty$. Equation (9) is strictly true only when square-law detectors are employed at the two receivers and would be only approximately true for linear detectors.

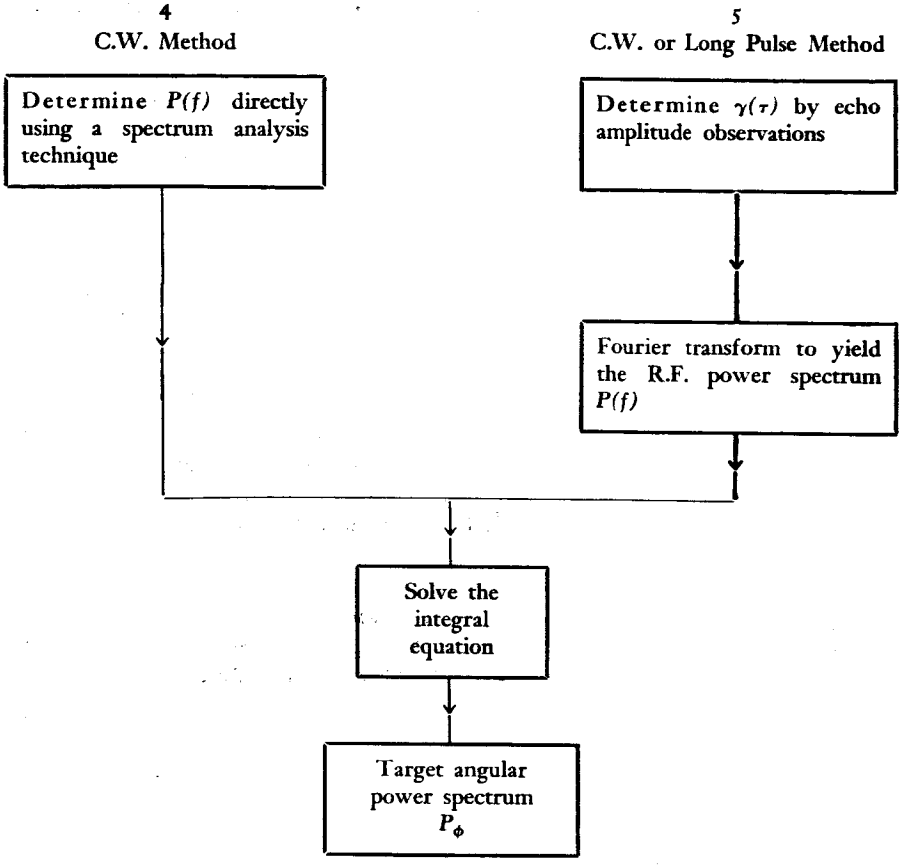
The function $\gamma(\Delta\omega)$ determines the bandwidth of a single channel communications system which could be built using the planet as a reflector. Ingalls *et al.*[17] determined $\gamma(\Delta\omega)$ and also $\gamma(\tau)$ for the Moon, and their results are shown in Figure 4. This three dimensional function $\gamma(\tau, \Delta\omega)$ is a double Fourier transform of the target scattering function $P(f, \tau)$ (Figure 1). Thus if the planet does not have a uniform surface and it is not possible to determine the target scattering function $P(f, \tau)$ (say because a pulse transmitter is not available), then instead of $\gamma(\tau, \Delta\omega)$ the double Fourier transform of $P(f\tau)$ must be determined.

III. Summary of Experimental Methods

When the target is a sphere having statistically uniform roughness but unknown rotation period there are three methods by which the angular power spectrum P_ϕ can be determined. These are shown schematically below:



If, however, the rotation period W is known then two further methods of determining the angular power spectrum P_ϕ may be employed.



In principle it is not necessary to know W to employ methods (4) and (5), since f_{max} can be determined by the same experiment and this yields W. In practice, however, little power is likely to be reflected from the limbs where few areas are normal to the line-of-sight. The signal-to-noise ratio for these reflections is, therefore, likely to be very poor and hence methods (4) and (5) are to be recommended only when W is known. As was stated earlier any pair of measurements which combine methods (1), (2), or (3) with either (4) or (5) will yield W. Alternatively W can be determined by spaced receiver cross-correlation measurements.

Finally, in methods (2), (3) and (5) it is necessary to use C.W. or long pulses. The length of the pulse T must be at least as great as the depth of the target $(2a/c)$. In method (4) the observed spectrum will be the con-

volution of the spectrum of the transmitted signals and $P(f)$. Consequently this method can only be employed when a C.W. radar is available.*

IV. The Surface Correlation Function $\rho(d)$

An irregular surface can best be described in terms of a correlation function which provides a measure of how the height h of the surface from some reference plane varies as a function of distance d measured over this reference plane. Consider a surface which has height fluctuations about the mean which are normally distributed with a standard deviation $h_m(= \sqrt{\overline{h^2}})$, and where the correlation coefficient $\rho(d)$ which relates these height variations to the horizontal distance in the plane follows a Gaussian law $\rho(d) = \exp(-d^2/2d_0^2)$. Here d_0 is the structure size being the value for d for which $\rho(d) = 0.61$. If a man were to stand at one point on this surface and then walk a distance d_0 in any direction the probability that he will then be the same height is 0.61. A Gaussian surface is an idealized surface which is rarely to be found in nature. However, because it is simple to treat mathematically, it has received a good deal of attention.

In the following sections the attempts of different workers to find a relationship between the surface correlation function $\rho(d)$ and the target angular power spectrum P_ϕ will be outlined. Most of this work was performed in order to discover the nature of the surface of the Moon from radar observations but is equally applicable to other distant spherical bodies.

IVa. Hargreaves' model

Hargreaves[13] was the first to attempt a statistical description of the surface features of the Moon. At the time he wrote his paper (1959) the best available law for the angular power spectrum was that published by Evans[7]

$$P_\phi = P_{(\phi = 0)} \cos^{30} \phi \qquad (10)$$

Hargreaves, therefore, attempted to find a model which would yield this law. This work was largely influenced by the studies of Ratcliffe[22] and Bowhill[1] who had developed suitable theory for the scattering of radio waves by irregularities in the ionosphere. Hargreaves proceeds as follows.

Consider a plane surface where the height fluctuations about the mean have a standard deviation h_m and the correlation function is of the form $\rho(d) = \exp(-d^2/2d_0^2)$. When plane waves of length λ are incident on this surface, the phase corrugations which appear in the reflected wavefront will have a standard deviation given by $\theta = 4\pi h_m/\lambda$ radians. The modulation is said to be shallow if $\theta < 1$ radian and the structure size l_0 in the reflected wave is then the same as d_0. (That is to say, the field correlation function $\rho_E(d)$ is closely related to the surface correlation function $\rho(d)$.) Where the modulation is deep ($\theta > 1$), a smaller structure size $l_0 = d_0/\theta$ is observed in the receding wave. Bowhill[1] has shown that at a great distance from the surface the initial phase corrugations of the re-

*Alternatively the phase-coherent pulse technique described by Pettingill[20] may be employed.

flected wave-front can be recognized either as amplitude or phase fluctuations, both of which have a mean value $\theta/\sqrt{2}$. Also the linear structure size in the wavefront will appear the same whether phase or amplitude variations are observed. The apparent angular power spectrum P_ϕ is the Fourier transform of the spatial correlation coefficient $\rho_E(d)$ which describes the electric field after reflection, and can be represented as $P_\phi = \exp(\phi^2/2\phi_0^2)$. Thus shallow modulation gives an angular power spectrum whose width is $\phi_0 = \lambda/2\pi d_0$ and for deep modulation $\phi_0 = \theta\lambda/2\pi_m d_0 = 2h/d_0$. These formulae are rigorously true for irregularities extending in one dimension only, but have been found numerically close to calculations for the two dimensional case.

Hargreaves next considers the lunar surface. He argues that because a very large fraction of the returned energy comes from the leading edge of the Moon, the approximation to a plane surface does not introduce large errors. The surface appears to impose deep modulation on the reflected signals. Hence the large irregularities distribute the power over $\phi_0 = 2h_m/d_0$. The result of Evans, $P_\phi \propto \cos^{30}\phi$, is nearly approximated by a Gaussian $P_\phi \propto \exp(-\phi^2/2\phi_0^2)$ where $\phi_0 = 0.18$ radian. Hence $h_m/d_0 \leqq 0.09$, and this means that the average surface gradient sampled in areas many wavelengths across is not steeper than 1 in 10.

It will be noted that when $h_m \geqq \lambda/4$, it appears to be impossible to determine both h_m and d_0 from measurements of the angular power spectrum P_ϕ. Instead only the ratio of these two quantities is obtained and this is a measure of the average gradient. Hargreaves was careful to point out that the above results apply only to large smooth areas on the surface. Small scale irregularities having vertical and horizontal dimensions of about one wavelength give rise to Lambert scattering, and at this time radio-echo studies of the Moon had not detected such reflections.

If it were necessary to determine separately the R.M.S. height variation h_m and the surface structure size d_0, it would seem that observations must be made at increasingly long wavelengths λ until the transition is reached at which the modulation changes from deep to shallow. The angular power spectrum would then be characterized by a width $\phi_0 = \lambda/2\pi d_0$, and by measuring this d_0 could be determined. It seems likely in the case of the Moon (and possibly some of the planets) that the wavelength required to reach this transition would be so long that it could not propagate through the earth's ionosphere.

IVb. Hagfors model

A more exact treatment of the reflection properties of sphere which has a Gaussian distribution of surface deviations has been given by Hagfors[12]. This author treats only the case where the deviations from a smooth sphere are much larger than the wavelength of the incident radio waves, and by using Huygen's principle to sum the reflected components he obtains for the angular power spectrum

$$P_\phi \cong P_{(\phi=0)} \exp\left|-d_0^2 \tan^2\phi/4h_m^2\right| \qquad (11)$$

The same result was obtained by Spetner and Katz[23] for a plane surface which has a superposition of specularly reflecting facets having a Gaussian distribution of slopes. We can compare this with Hargreaves' result.

$$P_\phi \cong P_{(\phi = 0)} \exp [- d_0^2 \phi^2/4h_m^2] \quad (12)$$

and see that since most of the power is returned for the region $O < \phi < 10^0$, the error in his approximate expression is not serious.

Hagfors has compared this result with the experimental results of Evans *et al.*[9] and finds that these yield a value for the average gradient $h_m/d_0 = 1$ in 15 whilst the results of Hey and Hughes[15] yield a value $h_m/d_0 = 1$ in 18. These two sets of experimental results were obtained at widely different wavelengths (Evans *et al.*, $\lambda = 3.0$ metres; Hey and Hughes, $\lambda = 10$ cms). The agreement between the two sets of results presumably indicates that the surface irregularities which cause the Moon to be nearly uniformly bright at visual and infra-red wavelengths must be many times smaller than the shorter wavelength.

Recently the work of Hughes[16] and Pettengill[20] has established that for the Moon the angular power spectrum is of the form

$$P_\phi = P_{(\phi = 0)} (\exp - 10.5 \sin \phi + \cos^{3/2} \phi/200) \quad (13)$$

The second term within the brackets predominates in the region $50° < \phi < 90°$ and this law may represent a combination of Lambert and Lommel-Seeliger scattering. At all events this term is attributed to the small scale features of the surface which have dimensions of the order of one metre. In the previous lecture, Dr. Pettengill concluded that for the Moon only some 5% of the surface is covered by such structure. The first term is similar to the law proposed by Evans[7] inasmuch as it accounts for the fact that the principal reflections from the Moon occur from a region at the center of the visible disc. However, the form of this law is clearly unlike a Gaussian $\exp (- \phi^2/2\phi_0^2)$. A comparison of this law $\exp (- 10 \sin \phi)$ with the Gaussian law which matches it best is shown in Figure 5. The failure of the Gaussian model of the surface to predict the observed angular power spectrum has led to a search for a better model.

IVc. Daniels' model

Daniels,[3-6] like Hargreaves, developed a solution based upon earlier work developed originally (by Feinstein[10]) for reflections from the irregular structure of the Earth's ionosphere. Daniels[6] shows that for an irregular surface extending in one dimension only that the relationship between the correlation function which describes the field E after reflection $\rho_E(d)$ and the surface $\rho(d)$ is

$$\rho_E(d) = \exp \{- \alpha[1 - \rho(d)]\} \quad (14)$$

where $\alpha = \overline{4k^2}\, h^2$, $k = 2\pi/\lambda$ is the wave number, λ is the radio wavelength, and $\overline{h^2}$ is the mean square amplitude of the height variations. Daniels assumes the autocorrelation function $\rho(d)$ describes a surface for which the

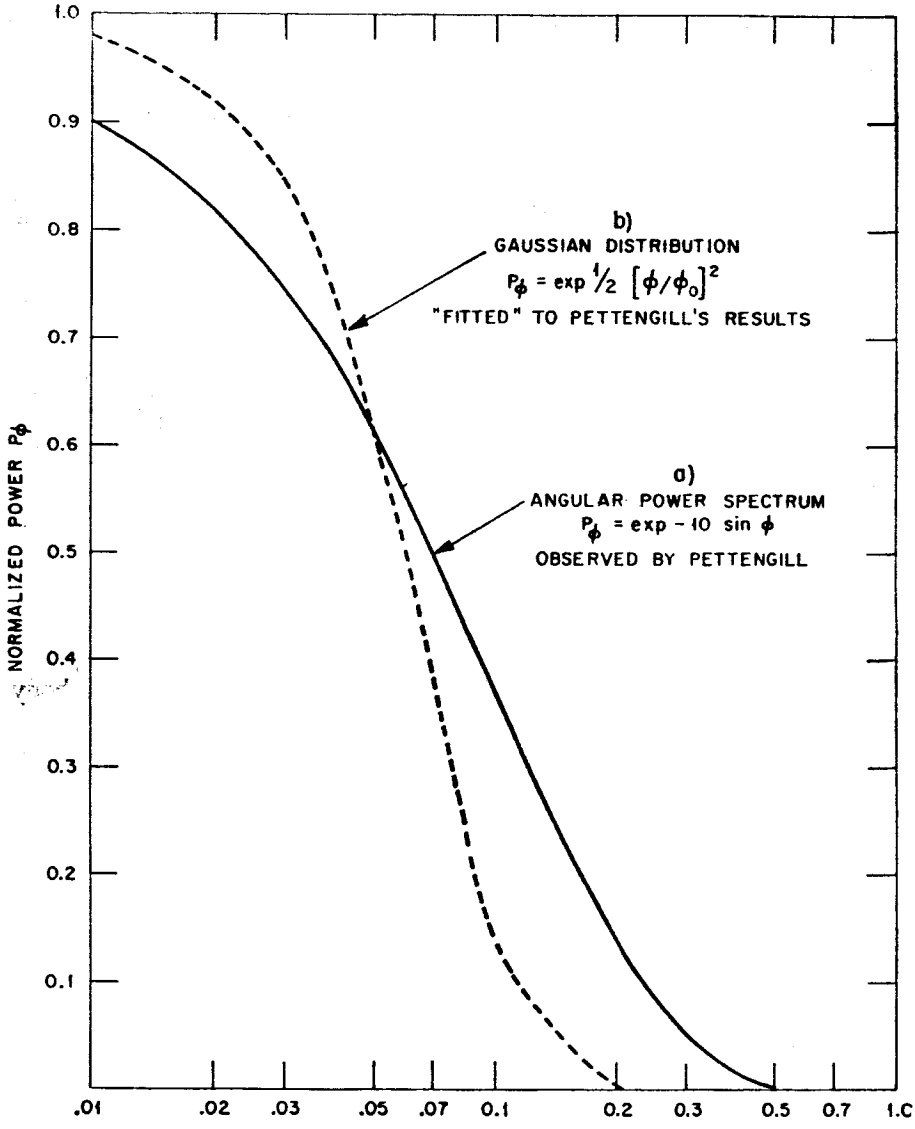


Fig. 5—Angular power spectra a) observed by Pettengill and b) the Gaussian distribution which best fits these results.

values of h, the height of the true surface from the mean are normally distributed. Further that the differences of the h values at all pairs of points x and x' ($x - x' = \mathrm{d}$) will also be normally distributed with a variance that is given by

$$\overline{(\Delta h)^2} = \overline{2h^2} \, [1 - \rho(\mathrm{d})] \tag{15}$$

Equation (14) holds only where α is large, i.e. $\sqrt{\overline{h^2}} >> \lambda/4$. Because the angular power spectrum P_ϕ is given by the Fourier transform of $\rho_E(d)$ it follows that

$$P_\phi = \text{Fourier transform of } \exp\{-\alpha\,[1 - \rho(d)]\} \tag{16}$$

This relationship provides a means by which the spatial correlation function which describes the surface $\rho(d)$ can be determined from measurements of the target angular power spectrum P_ϕ. When the surface is almost smooth ($\sqrt{\overline{h^2}} < \lambda/4$) then α becomes small and Equation (14) tends towards

$$\rho_E(d) \approx 1 - \alpha + \alpha\,\rho(d) \tag{17}$$

The Fourier transform of $\rho_E(d)$ would then be the sum of two terms: a delta function and a term proportional to the Fourier transform of $\rho(d)$. The delta function corresponds to the specular component which would be reflected by such a surface and the remaining term to the random component. Daniels[6] has also given an expression for the angular power spectrum of a two-dimensional curved surface

$$P(r) = \int_{-\infty}^{+\infty} dJ_0\,(2krd/a)\cos\,(kd^2/a)\exp\{-\alpha[1-\rho(d)]\}\,(d) \tag{18}$$

where r is the radial distance of a point on the disc from the center, $d =$ distance measured over the surface, $k = 2\pi/\lambda$, $a =$ the radius of the body and $\alpha = 4k^2\,\overline{h^2}$. (For small angles $r/a = \phi$)

In addition to investigating the relations between the target angular power spectrum P_ϕ and the surface correlation function $\rho(d)$, Daniels[3] has shown that $\rho(d)$ can be determined directly from measurements of the echo amplitude autocorrelation function $\gamma(\tau)$. A point on the surface defined by the intersection of the surface with a line drawn from the observer to the center of the body will fluctuate in height as the body rotates. If this function is defined by a correlation function $\Psi(\tau)$ then the autocorrelation function $R(\tau)$ observed for R.F. waveform at a single station on the Earth is given by

$$R(\tau) = \exp\{-\alpha\,[1 - \Psi(\tau)]\} \tag{19}$$

where $\alpha = 4k^2\,\overline{h^2}$, $k = 2\pi/\lambda$ and $\overline{h^2}$ is the mean square amplitude of the surface fluctuations. It follows from Eq. 19 that

$$\Psi(\tau) = 1 + (1/\alpha)\,\log\,R(\tau) \tag{20}$$

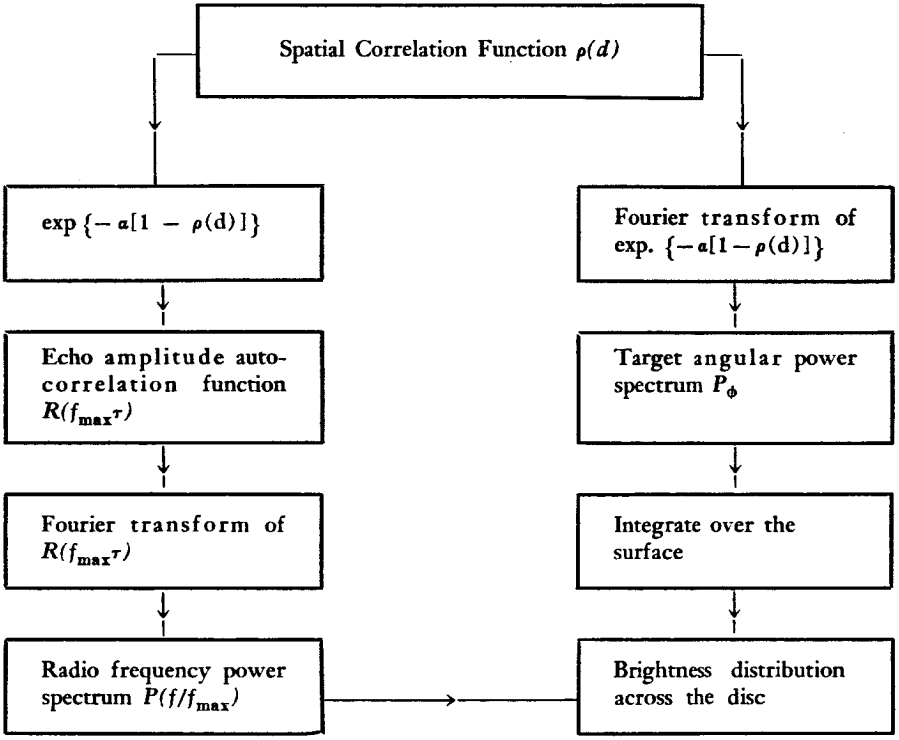
As stated earlier it is convenient to remove the effects of different rates of libration (for the Moon) by normalizing time τ by multiplying by the doppler shift f_{max} corresponding to reflection from one of the limbs. Now $f_{max} = 2aW/\lambda$ where a is the radius of the planet, and W is the rotation speed in radians/sec. Since $Wa\tau =$ distance d measured over the surface we have

$$\Psi(f_{max}\tau) = \rho(2d/\lambda) \tag{21}$$

Thus

$$\rho(2d/\lambda) = 1 + (1/\alpha) \log R(f_{max}\tau) \quad (22)$$

which gives the spatial correlation function $\rho(d)$ directly in terms of the observed variable τ. Daniels[3] also has shown how the spatial correlation function $\rho(d)$ can be determined directly by spaced receiver observations. These derivations employ Huygen's principle and no approximations are made of the planetary surface to a plane surface. The relationships between the spatial correlation function $\rho(d)$, the echo autocorrelation and the target angular power spectrum are summarized in the diagram below.



A similar diagram can be drawn for the relationships where spaced receiver measurements are made of the echo amplitude correlation function (discussed in (2b), since this is an alternative method by which the brightness distribution across the disc can be determined.

The echo autocorrelation results (for the Moon) of Evans *et al.*[9] and Ingalls *et al.*[17] have been presented by the author[8] as shown in Figure 6. Daniels[6] has argued that these results can be approximately represented by a simple function

$$\gamma(f_{max}\tau) = \exp(-f_{max}\tau) \quad (23)$$

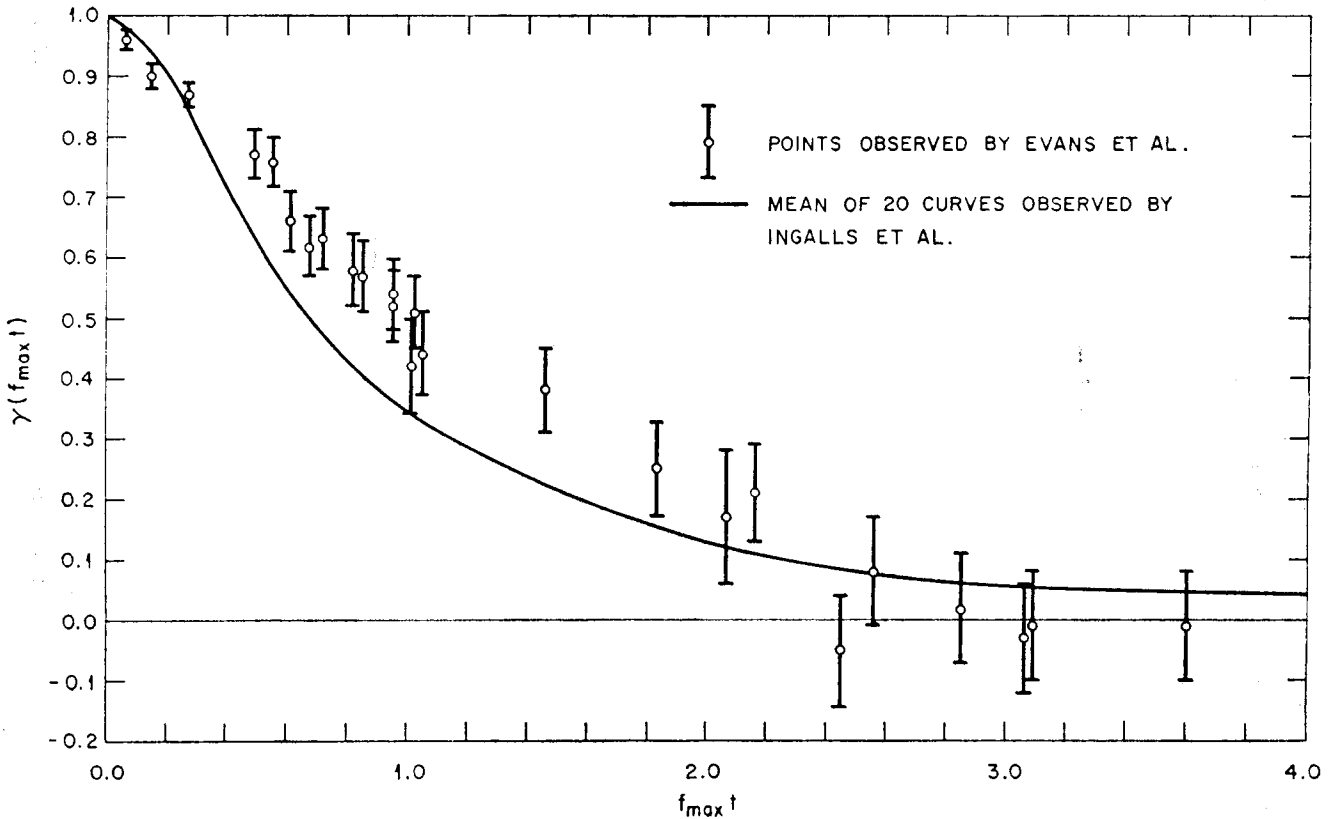


Fig. 6—The autocorrelation functions of echo amplitude $\gamma(f_{max}\tau)$ observed by Evans et al.[9] and Ingalls et al.[17]

Since the correlation function of the R.F. waveform $R(f_{max}\tau)$ is related to $\gamma(f_{max}\tau)$ approximately in (See Section Ic)

$$\gamma(f_{max}\tau) \approx [R(f_{max}\tau)]^2 \tag{24}$$

it follows that $R(f_{max}\tau) \approx \exp(-f_{max}\tau/2)$ (25)

and the Equation (22) becomes $\rho(2d/\lambda) \approx 1 - f_{max}\tau/2\alpha$

$$\approx 1 - d/\alpha\lambda \tag{26}$$

Because $\gamma(f_{max}\tau)$ tends rapidly to zero (Fig. 6) it can be seen that α, which is the asymptotic value of $-\log \gamma(f_{max}\tau)$ cannot readily be determined. In other words, the experimentally determined form of $\gamma(f_{max}\tau)$ for the Moon has significance only over small values of $f_{max}\tau$, which correspond to small values of d. Hence Daniels concludes that in practice it is not possible to determine the scale of the surface roughness $\sqrt{\overline{h^2}}$ from radar fading measurements. Daniels[6] has, however, attempted to determine a value for α from the contour map of the limb region of the Moon which was constructed by Hayn[14] from photographs of the lunar profile. That is, Daniels determined the correlation function $\rho(d)$ for the large-scale features of the Moon which are just visible optically. The results of this work are shown in Fig. 7. This process may be criticized on the ground that prominent features nearer to the Earth than the limb may cause some shadowing thereby giving rise to a false determination of $\rho(d)$. By fitting a Gaussian curve to his results, Daniels concludes that the scale of coherence d_0 is about 90 Kms and the mean square height fluctuation $\overline{h^2} = 1.85 \times 10^6$ m^2. At the wavelength used by Ingalls[17] ($\lambda = 68$ cms) the corresponding value of α is 6.3×10^8. The correlation function corresponding to the radar results

$$\rho(2d/\lambda) = 1 - d/\alpha\lambda \tag{27}$$

is plotted in Figure 7 for this value of α. This is shown as the uppermost curve, and it can be seen that there is no agreement with the optical results. Daniels concludes that this demonstrates that the portion of $\rho(d)$ deducible from the radar data alone is a small portion near the origin, from which only the initial slope can be determined. If now the correlation function $\rho(d)$ is written in a form which approximates to the radar results for small d, and to the optical results for large d we have

$$\rho(d) = (1 - d/\alpha\lambda) \exp(-d^2/d_0^2) \tag{28}$$

where $d_0 = 90$ km, $\lambda = 68$ cms, and $\alpha = 6.3 \times 10^8$.

Daniels has substituted this expression into the function for the target angular power spectrum of a two dimensional surface, Equation (18), and evaluated the integral with a digital computer. The power spectrum $P(r/a)$ ($r/a = \phi$ for small angles) is shown in Figure 8, where Pettengill's result $P_\phi = \exp(-10\, r/a)$, is also plotted. The agreement seems fairly good. If the integral is reevaluated omitting from the expression for $\rho(d)$, Equation

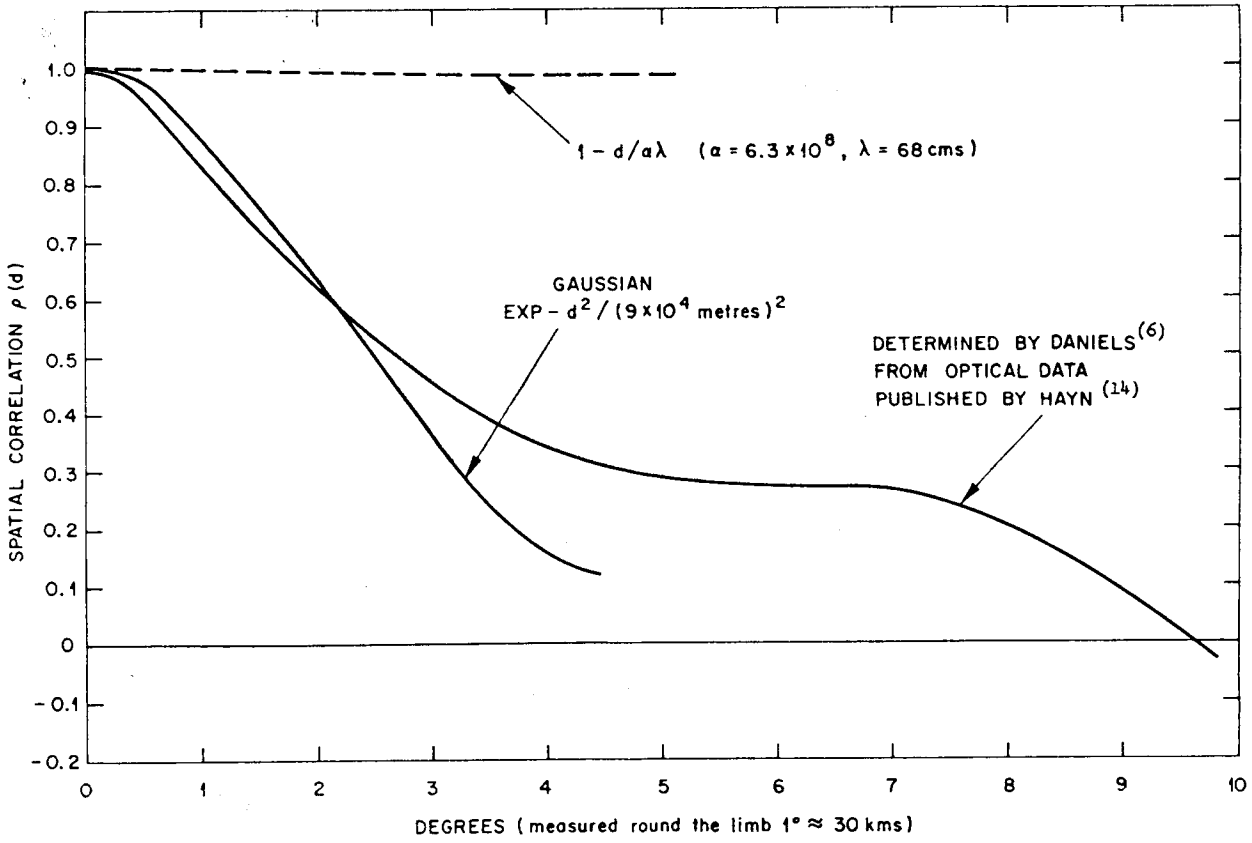


Fig. 7—The spatial correlation function for the lunar surface determined by Daniels.[6]

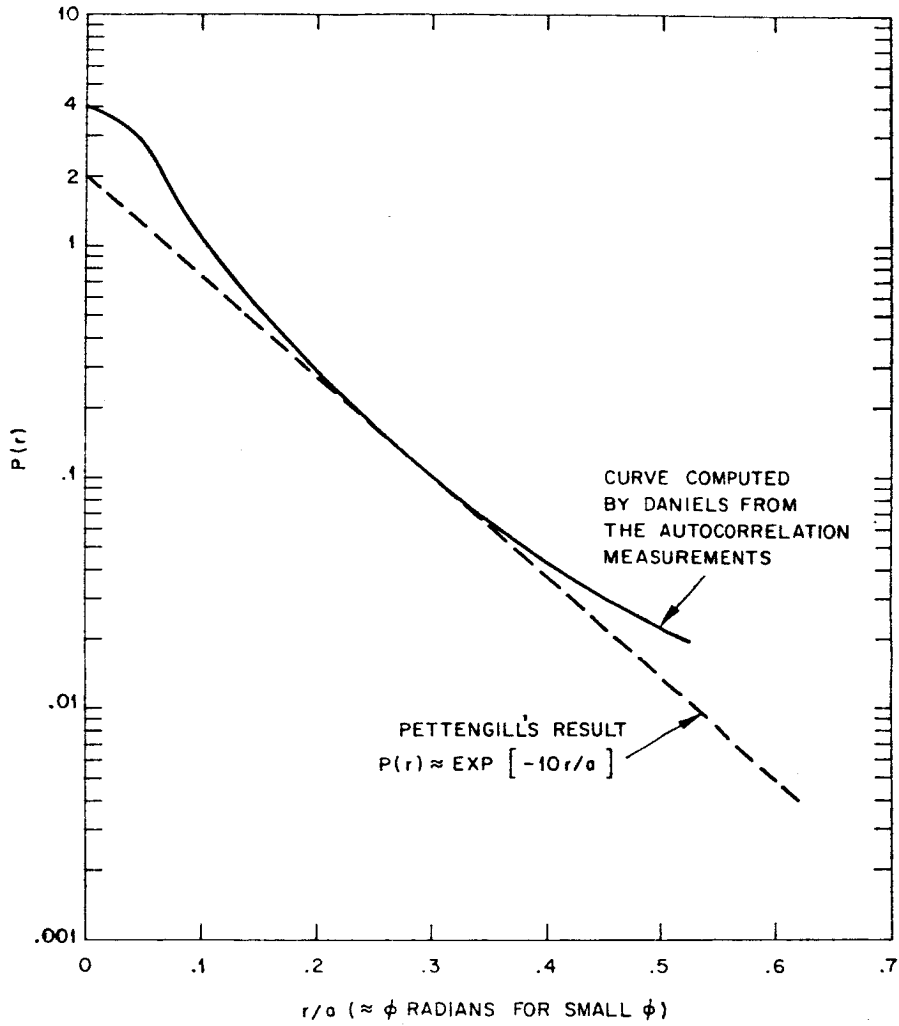


Fig. 8—The angular power spectrum.

(28), the Gaussian term corresponding to the optical results, an almost identical curve for the angular power spectrum P_ϕ is obtained. Thus Daniels concludes that the angular power spectrum observed by Pettengill, $P_\phi \propto \exp(-10\,\phi)$, is due entirely to surface irregularities of the order of tens of wavelengths across. It seems therefore that though this work is better able to predict the angular power spectrum it offers little hope that by radar measurements alone at one frequency the complete form of the spatial autocorrelation function can be determined.

V. Summary

The experimentally observed angular power spectrum for the Moon $P_\phi = \exp(-10 \sin \phi)$ is distinctly different from that to be expected from a surface having a Gaussian distribution of surface slopes.

$$P_\phi \propto \exp - [\phi/\phi_0]^2$$

The most extensive treatment of the problem seems to have been that by Daniels[6] who has succeeded in showing that both the angular power spectrum measurements $\{P_\phi\}$ and the echo autocorrelation measurements $\{\gamma(f_{max}\tau)\}$ yield a similar result for the spatial correlation function for the lunar surface $\rho(d)$. Daniels' treatment assumes that the height variations h of the surface are normally distributed about the mean, and that these are described by a correlation function $\rho(d)$ where

$$\rho(d) = \frac{\overline{h(x)\ h(x+d)}}{\overline{h^2}}$$

Then from the published echo autocorrelation measurements[8], Daniels has shown that for the Moon

$$\rho(2d/\lambda) \approx 1 - d/\alpha\lambda \quad (\alpha = 4k^2\ \overline{h^2})$$

The value of α cannot be determined from the radar measurements alone, because of the large value of $\overline{h^2}$ ($= 1.85 \times 10^6$ m^2). Thus the radar results are relevant to a small part of the correlation function $\rho(d)$ near the origin (d, small), and yield no information about the large scale structure which is visible optically. In short, radar measurements yield information only about the presence of irregularities on the surface having sizes in the range just less than one to many tens of wavelengths. Structure which is considerably smaller than the wavelength may never be detected, and large features on the surface may only be examined if they are themselves not covered by irregularities of an intermediate size. Where this is the case (as for the Moon) the radar results yield information principally about those elements of the surface whose structure size is not very different from the wavelength.

Thus in the case of the Moon, we may conclude from the radar results that 5% of the surface is covered with irregular structure (perhaps boulders) of the order of 1 metre in size. The remainder of the surface appears to be smooth and undulating with an average gradient in the region of less than 1 in 10. The horizontal or vertical scales of these undulations has not been determined, and there seems no simple improvement in either the experimental tchniques or the theory which will yield this information. Whilst the available evidence is still meagre it appears that Venus scatters in a very similar manner to the Moon. At optical and infra-red wavelengths, the Moon appears to be uniformly bright. Thus it follows that the surface must be rough at wavelengths of the order of 10^{-2} cms. It is not known at what wavelength between $\lambda = 10$ cms and 10^{-2} cms the transition occurs between

quasi-smooth scattering and uniformly bright scattering. It seems unlikely that this information could be obtained by Earth-based radar equipment in view of the strong atmospheric absorption in this wavelength range.

VI. Conclusion

Radar methods of studying distant planetary objects can yield but a crude picture of their surfaces. By measuring the gross reflectivity some idea of the kind of surface (water, or land) can be obtained. Other radar studies which have been described in this lecture provide a means of determining the gross roughness characteristics of the surface. At present it seems that such measurements only provide a measure of the fractional amount of surface which has a roughness of the same size as the wavelength, and the average gradient of the remainder. Although radar measurements of the Moon have been made over a frequency range which spans some six octaves (15 meters to 3 cms wavelength), there seem to be no marked changes in the scattering characteristics as a function of the frequency of the exploring radio wave. It seems that observations substantially beyond these limits are required but these must await the development of satellite-borne radar equipment which will be unhindered by the absorption of the Earth's atmosphere.

References

1. Bowhill, S. A., J. Atmos. Terr. Phys., *11*, 91, 1957.
2. Bracewell, R. N., Proc. IRE *46*, 97, 1958.
3. Daniels, F. B., USASRDL Technical Report 2110, 1960.
4. Daniels, F. B., USASRDL Technical Report 2163, 1960.
5. Daniels, F. B., Nature *187*, 399, 1960.
6. Daniels, F. B., J. Geophys. Res. *66*, 1781, 1961.
7. Evans, J. V., Proc. Phys. Soc. *B 70*, 1105, 1957.
8. Evans, J. V., Lincoln Laboratory M.I.T. Group Report 3G-0004.
9. Evans, J. V., Evans, S., and Thomson, J. H., "I.A.U. Paris Symposium on Radio Astronomy," edited by R. N. Bracewell, p. 8 (Stanford: Stanford University Press) 1959.
10. Feinstein, J., Transactions of the IRE *AP-2*, 23, 1954.
11. Fricker, S. J., Ingalls, R. P., Mason, W. C., Stone, M. L., and Swift, D. W., Lincoln Laboratory M.I.T. Technical Report 187.
12. Hagfors, T., Stanford Electronics Labs. Scientific Report, No. 8, 1960.
13. Hargreaves, J. K., Proc. Phys. Soc. B., *73*, 536, 1959.
14. Hayn, F., Sächsiche Ak. Wiss. Adh. Math. Phys., K. 1, *33*, 5, 1917.
15. Hey, J. S. and Hughes, V. A., "I.A.U. Paris Symposium on Radio Astronomy" edited by R. N. Bracewell, p. 13 (Stanford: Stanford University Press), 1959.
16. Hughes, V. A., Nature *186*, 873, 1960.
17. Ingals, R. P., Bird, L. E., and Day, J. W. B., Proc. IRE *49*, 631, 1961.
18. Leadabrand, R. L., Dyce, R. B., Fredriksen, A., Presnell, R. I., and Schlobohm, J. C., Proc. IRE *48*, 932, 1960.
19. Manasse, R., Lincoln Laboratory M.I.T. Group Report 312-23, 1959.
20. Pettengill, G. H., Proc. IRE *48*, 933, 1960.
21. Price, R., Privately communicated, 1960.
22. Ratcliffe, J. A., Rep. Prog. Phys. *19*, 188 (London: Physical Society) 1956.
23. Spetner, L. M. and Katz, L., Transactions of the IRE *AP-8*, 242, 1960.

RADAR MEASUREMENTS IN THE SOLAR SYSTEM

G. H. PETTENGILL
Lincoln Laboratory, Massachusetts Institute of Technology*

I. Introduction

Dramatic announcements in recent years have served notice that radar is now capable of receiving sensible echo power from the nearer planets and the sun. What can be learned through the application of radar to studies of the solar system? Radar is unique in providing the investigator with a measure of control over the outgoing, illuminating radiation. By studying the changes induced in the echo power received back, he may hope to gain information concerning: (1) the distance to the echoing region, (2) the motion of the echoing region along the observer's line-of-sight, (3) some characteristics of the surface roughness and composition, (4) the magnitude and orientation of the planetary axial rotation, (5) the existence of a planetary magnetic field and ionosphere, and (6) the amount and state of motion of interplanetary plasma. (The ranking of the various objectives is approximately as the relative difficulty.) Measurements of (1) and (2) will be discussed briefly in a later section of this presentation; (3) and (4) form the subject of "Radar Studies of Planetary-Surfaces." Objectives (5) and (6) have not yet been explored to any large extent and will not be discussed in these lectures.

Figure 1 arrays the potential targets which exist in the solar system according to their detectability. The reference chosen is the moon, and it should be noted that once the 68-db gap separating the moon from Venus is closed a large number of targets become available with only a modest further increase in radar performance. In plotting the data of Fig. 1 it has been assumed that the reflectivity in each case is the same as for the moon and that the echo strength obtained will vary as the square of the radius of the target and as the fourth power of its distance. Further it is assumed that the rotation of the target and the associated spectral broadening will play no part in degrading the signal detectability. It is shown in the lecture on Planetary Surfaces, Part I, that while the assumptions for reflectivity may be approximately valid, at least for the smaller planets and moons, the rotation rate may considerably change the relative detectability when radars capable of observing with extremely narrow spectral resolution are used. The sun has been specifically excluded from Fig. 1, because it presents special problems as a reflector of radar energy.

*Operated with support from the U. S. Army, Navy and Air Force.

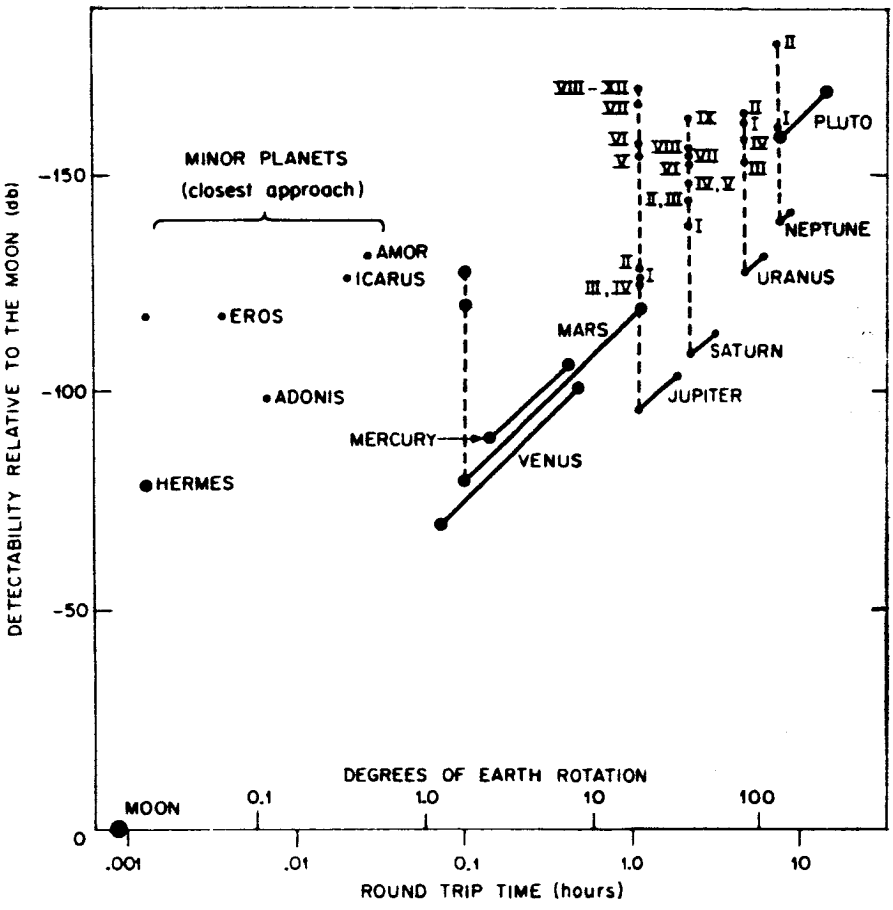


Fig. 1—Relative system sensitivity required to develop a given signal-to-noise ratio from various bodies ignoring surface conditions. Roman numerals refer to satellites.

Table I lists the relative capability against "hard" targets of large radar systems now in existence or shortly to be completed. A "hard" target is defined as one which possesses a relatively dense dielectric or conducting surface presenting an abrupt discontinuity to the impinging electromagnetic wave. By contrast, the sun (and possibly the giant planets) are considered "soft" targets since considerable penetration of their constituents occurs before reflection can take place.

A further distinction is made between targets like the moon, Mercury and Venus which are rotating slowly and thus do not greatly broaden the spectrum of the returned energy, and Mars which, like the earth, is rotating sufficiently to significantly disperse the reflected frequency spectrum. When it is noted that the present Millstone was able to observe echoes from Venus with about a 13-db signal-to-noise ratio during the close approach of April,

TABLE I

Approximate Relative Sensitivity of Contemporary Radars

Radar System	Sponsor	Completion Date	Freq. (Mcps)	Relative Capability Against Rapidly Rotating Targets	Relative Capability Against Slowly Rotating Targets
Millstone	M.I.T.	Now	440	1	(100)
Jodrell Bank	U. of Manch.	Now	408	2	20
Goldstone	J.P.L.	Now	2388	0.2	10
El Campo	M.I.T.	Now	38	1	(100)
Arecibo	Cornell	1962	430	20,000	(2,000,000)
Millstone	M.I.T.	1962	1295	20	(500)
Haystack	M.I.T.	1963	(8000)	200	200

The parentheses in the right-hand column, above, indicate that the improved capability rests on digital processing techniques carried out at some time following the collection of data.

1961, it will be seen by referring to Fig. 1 that a large number of the closer planets should become accessible to radar investigation over a broad range of frequencies within the next several years.

II. The Determination of the Astronomical Unit

As the brightest celestial object after the sun and the moon, the planet Venus has drawn attention since ancient times. Its angular position in the sky has been measured with precision for centuries, and from the accumulated body of optical measurements a highly precise orbit has been formulated[1, 2]. This orbit, however, in common with the orbits of all other members of the solar system (but specifically excluding the earth-moon system) is expressed accurately only in terms of an exterior unit of length: the Astronomical Unit. This unit is defined as the mean distance from the sun to the earth in its orbit, and may be related to terrestrial standards, such as the meter, with only mediocre accuracy. Prior to 1961 there were a number of reputable determinations of the A.U. in the literature ranging from approximately 149,470,000 to 149,670,000 kilometers[3], a variation about the mean of about one part in 1500. By contrast the internal inconsistencies between the relative positions of most of the members of the solar system are probably less than one part in 150,000.

An alternate way of expressing the A.U., frequently met with in celestial mechanics, is as the equatorial horizontal parallax of the sun, or Solar Parallax. The Solar Parallax is defined as the angle which is subtended by the equatorial radius of the earth at a distance of one A.U. and is approximately equal to 8.8 seconds of arc.

Because of the great accuracy connecting the positions of the planets to the A.U., a single accurate measurement of the distance in kilometers to any of the inner 8 planets at a known time may be used to calibrate the A.U.

in terms of terrestrial units. In a related way, a measurement of planetary velocity with respect to the earth may also be used to express the A.U. in terrestrial units. In recent months, several radar systems have succeeded in measuring both the distance and velocity of Venus in its orbit, and have obtained from these results comparatively precise values for the A.U. Table II summarizes these findings.

TABLE II

Comparison of Radar Determinations of the A.U.

	Astronomical Unit (km)	Solar Parallax (secs of arc)
MIT Lincoln Laboratory, Millstone[4]	149,597,700 ± 1200	8.79450 ± 0.00008
Jet Propulsion Laboratory, Goldstone[5]	149,599,000 ± 1500	8.79444 ± 0.0001
Univ. of Manchester, Jodrell Bank[6]	149,601,000 ± 5000	8.7943 ± 0.0003

Both the Millstone and Goldstone radar measurements were made to a basic accuracy which in some cases approached one part in 10^6. The errors indicated above, therefore, for the determinations made by these two groups were largely set by inaccuracies in the planetary orbits of the earth and Venus. An estimate of the magnitude of these errors as well as an indication of the way in which they appear may be seen in Fig. 2. The differences between the observed distance to the planet Venus and the value obtained by computation from the ephemeris[1] are plotted for the approximately two-month interval over which observations were made at Millstone. The computed values were obtained as described in Ref. 7. A systematic variation in the range discrepancies are quite evident over the observing interval. Note that no simple adjustment in the value assumed for the Solar Parallax, π, will remove the systematic effect. Interestingly enough, if one makes a simple time displacement for the location of Venus in its orbit, a curve is obtained which fits the observed data very well.

Attempts are currently underway to compute the effects of a correction to the elements of Venus proposed by Duncombe[2]. No matter what the outcome of these calculations, it seems clear that accurate radar measurements of the planetary distances and velocities will eventually permit substantial improvement in the knowledge of planetary orbits.

III. Solar Radar Measurements

The detection of radar echoes from the sun involves a somewhat more complicated consideration than in the case of the moon and planets since the sun is surrounded by a dense, hot plasma. There is no simple surface presenting an abrupt discontinuity. Rather, the incident radio waves will encounter an increasingly dense plasma as they penetrate into the solar corona, until they reach a region whose plasma frequency equals that of the radiation. During the traversal of the outer regions of the corona, the radio waves will undergo absorption.

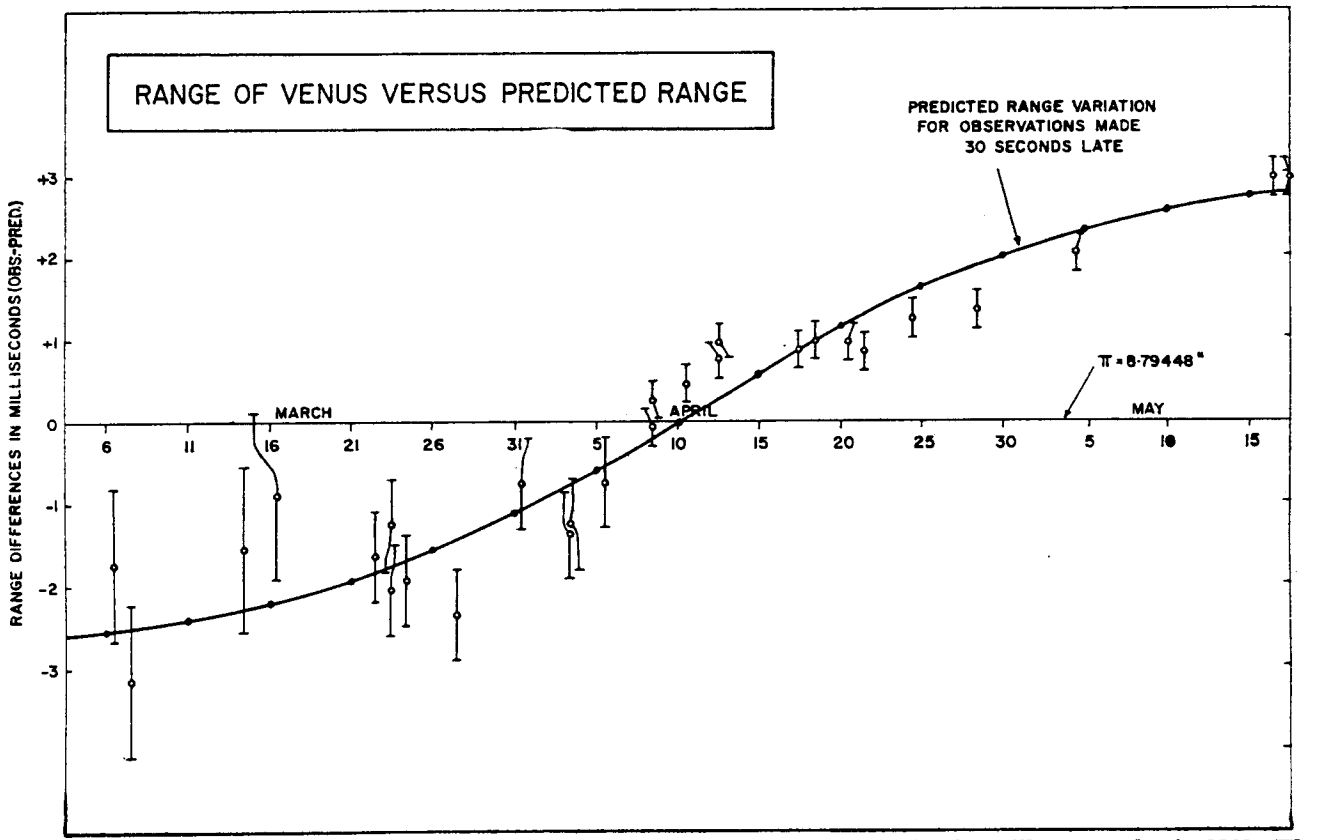


Fig. 2—Comparison between the observed and predicted distances to Venus obtained by the Millstone Radar in 1961. The predictions were based on Newcomb's elements for Venus as tabulated in Ref. 8. The solid curve shows the effect of a 30-second displacement in the times assigned to the measurements.

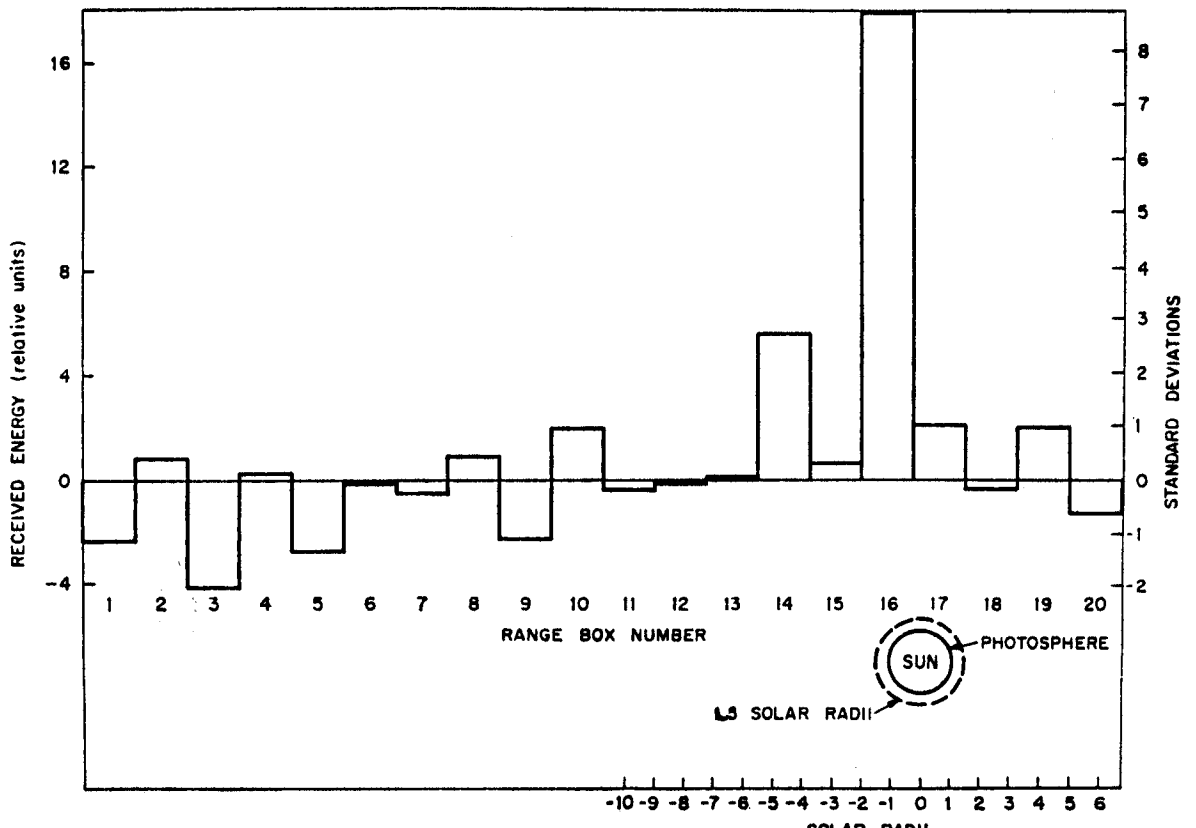


Fig. 3—Summation of the energy received in each range box relative to the position of the sun for the period April 19 to July 7, 1961.

The signal energy which is returned to the terrestrial receiving station must compete with the intense noise emission of the solar corona itself. Since the noise power emitted by the sun varies with frequency, reaching higher temperatures at lower frequencies, while the attenuation in the solar corona suffered by the entering and emerging radar energy tends to decrease at lower frequencies, one expects that an optimum frequency should exist for carrying out radar measurements. The first calculations to determine this frequency and to estimate the echo strength to be expected were carried out by Kerr[8]. Using a simplified model for the solar corona, Kerr finds an optimum frequency (for his assumed radar) of 30 Mcps with an associated coronal attenuation of 3.5 db. Some years later a paper by Bass and Braude[9] repeated the calculation for somewhat different assumptions and concluded that the optimum frequency should lie close to 50 Mcps. While these two results disagree, they both indicate that solar echoes may most readily be obtained at frequencies low in the VHF spectrum.

The first echoes from the sun were obtained by Von Eshleman[10] in 1959 using a frequency of 26 Mcps. These measurements yielded a cross section approximately in agreement with that predicted by Kerr[8] lying at a range corresponding to 1.5 solar photospheric radii. Because of the fixed antenna employed, observations were possible for only a few days.

More recently Abel[11] has obtained echoes using a partially steerable array at 38 Mcps. These data taken over several months' time show a cross section greatly below that observed earlier at 26 Mcps which varies greatly from day to day. Figure 3 gives a summation of results obtained by Abel et al.[11] over a 2½ month period at El Campo, Texas. Note that the returned energy lies in a single integration box at the expected range.

The variability observed in the recent experiments was unexpected and as yet is not understood. There has been at least one case where an enhanced return was associated with a period of increased solar activity. Obviously the solar conditions are much more complex than assumed by Kerr; perhaps with the addition of measurements over a range of frequencies and solar conditions, at least some partial understanding will emerge.

References

1. Astr. Papers of the Amer. Ephemeris, Vol. 15 (Washington, USGPO, 1955).
2. Astr. Papers of the Amer. Ephemeris, Vol. 16 (Washington, USGPO, 1956).
3. McGuire, J., Spangler, E. R., Wong, L., Sci. American, *204*, 64 (1961).
4. The Staff, Millstone Radar Observatory, Nature, *190*, 592 (1961).
5. Victor, W. K., and Stevens, R., Science, *134*, 46 (1961).
6. Thomson, J. H., et al., Nature, *190*, 520 (1961).
7. Pettengill, G., and Price, R., Planet. Space Sci., *5*, 70 (1961).
8. Kerr, F. J., Proc. I.R.E., *40*, 660 (1952).
9. Bass, F. G., and Braude, S. I. A., Ukr. Jour. of Phys., *2*, 149 (1957).
10. Eshleman, V. R., et. al., Science, *131*, 329 (1960).
11. Abel, W. G., et al., Jour. of Geo. Res., to be published (1961).